

ISOMORPHISM OF PITCH AND TIME

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Xin Wen

May 2019

© 2019 Xin Wen

ISOMORPHISM OF PITCH AND TIME

Xin Wen, Ph. D.

Cornell University 2019

Humans listen to music in daily life through effortlessly processing rich and complex auditory stimuli. Music cognition research taps into underlying principles in perception by studying shared musical preferences across various training backgrounds. This dissertation focuses on the two most fundamental dimensions of music, pitch and time, and explores the isomorphism in their perceptual preferences through six experiments. Two sets of structurally isomorphic patterns are used. The first set is the standard rhythmic pattern and the diatonic scale pattern (Chapters 2-4). The second set is the bell rhythmic pattern and the pentatonic scale pattern (Chapters 5-7). Chapter 1 introduces basic musical concepts about pitch, time, and their interaction. Chapter 8 compares the two sets of patterns in terms of cognitive principles underlying experimental results and presents future directions.

The seven standard rhythms and the seven diatonic scales have the surface structure of 2 2 1 2 2 2. Experiment 1 (Chapter 2) devised a novel probe accent technique to measure the rhythmic stability of each event in each standard rhythm. Experiment 2 (Chapter 3) used the probe tone technique to measure the tonal stability of each tone in each diatonic mode. Experiment 3 (Chapter 4) measured the fit judgment between each standard rhythm and each diatonic mode. Combinations were preferred in which tonally stable tones (governed by early occurrence, proximity to C

on both the chroma circle and the circle of fifths, and tonality) occurred in temporal positions where accents were preferred (governed by subgrouping and a syncopation-shifted metrical hierarchy).

Related to the first set of patterns, the five bell rhythms and the five pentatonic scales have the surface structure of 2 2 3 2 3. Experiments 4 (Chapter 5), 5 (Chapter 6), and 6 (chapter 7) used the same experimental designs to measure the rhythmic stability of each bell rhythm, the tonal stability of each pentatonic mode, and the fit judgement of rhythm-mode combinations, respectively. Similarly to the first set, combinations were preferred in which salient tones (governed by early occurrence, large interval size, and consonance) occurred in salient temporal positions (governed by early occurrence and long tone duration).

BIOGRAPHICAL SKETCH

Xin Wen was born and grew up in Beijing, China, where she received rigorous education. At a young age, she enjoyed singing and dancing, played the piano and switched to the flute at age 7. She adapted the name Olivia before attending the United World College of Adriatic, an international high school in Duino, Italy. The friends she made and the frequent travelling set the tone for her adult life.

Olivia pursued her passion in Psychology and Music in college and received her bachelor's degree from Grinnell College in Grinnell, Iowa, in the United States. She was actively involved in research in a developmental psychology lab at Grinnell College, and helped at a social cognition lab during her study-abroad semester at University College London in the United Kingdom. Thanks to generous scholarship, Olivia picked up multiple musical instruments such as guitar, harp, organ, marimba, and steel pan. Upon graduation, she held a flute recital, recorded a CD album singing Chinese folk songs while playing guitar, and choreographed and hosted a Salsa dance show. After graduation, Olivia spent a short year working as the lab manager in a psycholinguistics lab at University of Notre Dame in South Bend, Indiana.

In graduate school at Cornell University in Ithaca, New York, Olivia studied various topics in the field of music perception and cognition. These include repetition and interest, pitch and time parallel and interaction, and musical groove and social affiliation. She also explored research in dance at a motion-capture lab at University of Jyväskylä in Jyväskylä, Finland. As a result, she finds herself dancing Argentine tango on a weekly basis.

ACKNOWLEDGMENTS

First and foremost, I am grateful to my loving parents, Lijun Qian and Liang Wen, who provided constant emotional and mental support on a weekly basis over the 12 years that I have been away from home. They celebrated moments of my success and fueled me with comfort and encouragement at difficult times. I could not have made it this far without them.

My utmost gratitude goes to my Ph.D. advisor, Carol Krumhansl. As a mentor, she guided and nurtured me for the past 5 years. Her thoroughness from big pictures to tiny details held me to a high standard. As a collaborator, she listened keenly with a modest attitude and provided succinct comments that inspired me profoundly. As a role model, her diligence and persistence stimulated me during our 2-hour weekly meetings and motivated me to be productive. My academic thinking is strongly influenced by her intellect and I am grateful for this journey.

I am in dept to two collaborators who personally invested in me and financially supported my research. Vivian Zayas introduced to me ways of thinking in social psychology and equipped me with specific skills in designing questionnaires, coding, and data analysis. I treasure our conversations about dance and will always remember the excitement in her eyes when she talks about research ideas and statistics. Petri Toiviainen provided ample resources for me to learn the motion-capture technology and analyze motion data for two summers in Finland. I thank him, my collaborators in his team, Birgitta Burger and Joshua Bamford, and other fellow researchers in Jyväskylä. They were brilliantly reassuring and made everything simple and fun.

Many others have helped, supported, and cared for me during my graduate career.

I thank other members of my special committee, James Cutting and Katie Kinzler, for their constant support. I thank David Field for his honest comments on my presentations which I found extremely helpful. My thank also goes to the staff members in the Department of Psychology for their assistance, especially Pamela Cunningham whom I formed a special bond with. Last but not the least, I am grateful to my dear friends here in Ithaca. I feel especially fortunate to be close to Xi Shen, whose unconditional care accompanied me through the most difficult times in life.

Thank you to everyone who contributed to or witnessed my growth, and to all funding sources that have supported my education and career in various countries. I look forward to joining Molly Henry's research group as a postdoctoral researcher at the Max Planck Institute for Empirical Aesthetics in Frankfurt, Germany.

To my mother, Lijun Qian, and my father, Liang Wen.

TABLE OF CONTENTS

Chapter 1	Introduction	1
	Rhythm and Metrical Hierarchy	2
	Scale and Tonal Hierarchy	4
	Pitch-Time Interaction	9
Chapter 2	Isomorphism in the Standard and Diatonic Patterns	11
	Experiment 1	17
	Objectives of the Study	25
	Method	28
	Results	29
	Discussion	36
Chapter 3	Experiment 2	39
	Objectives of the Study	43
	Method	45
	Results	46
	Discussion	51
Chapter 4	Experiment 3	53
	Objectives of the Study	55
	Method	56
	Results	57
	Discussion	60
	Discussion for Experiments 1, 2, and 3	61

Chapter 5	Isomorphism in the Bell and Pentatonic Patterns	66
	Experiment 4a&b	70
	Objectives of the Study	73
	Method for Experiment 4a	73
	Method for Experiment 4b	75
	Results	76
	Discussion	79
Chapter 6	Experiment 5a&b	80
	Objectives of the Study	81
	Method for Experiment 5a	81
	Method for Experiment 5b	82
	Results	83
	Discussion	87
Chapter 7	Experiment 6	88
	Objectives of the Study	88
	Method	88
	Results	89
	Discussion	92
	Discussion for Experiments 4, 5, and 6	93
Chapter 8	General Discussion	94
	Comparison between the Standard and Diatonic Patterns, and the Bell and Pentatonic Patterns	94
	Subset Relation	96
	Complementary Relation	99
References		103

LIST OF FIGURES

1.1. Metrical hierarchy (Lerdahl & Jackendoff, 1985; Temperley, 1999 with modifications).	3
1.2. The chroma circle.	5
1.3. Circle of fifths with major keys in upper case letters and minor keys in lower case letters.	6
1.4. Probe tone profile for C major and C minor keys (Krumhansl & Kessler, 1982).	8
2.1. Theoretical structure of the standard pattern in clock position. The black circles mark temporal events that sounded in the standard pattern. Starting from clock position 0 going clockwise, the lines connecting consecutive black circles form durations of 2 2 1 2 2 2 1 (redrawn from Toussaint, 2013, Figure 14.8).	12
2.2. Theoretical structure of the diatonic pattern in tone identity. The black circles mark pitches that sounded in the diatonic pattern. Starting from the pitch C going clockwise, the lines connecting consecutive black circles form semitone intervals of 2 2 1 2 2 2 1 (redrawn from Toussaint, 2013, Figure 14.8, with modifications).	13
2.3. The seven unique rhythms of the standard pattern in clock position. Each of the seven rotations starts with a successive duration. For example, Rhythm 2 starts with the second duration in Rhythm 1 (redrawn from Toussaint, 2013, Figure 14.8, with modifications).	14
2.4. The seven unique modes of the diatonic pattern in tone identity. Each of	

the seven rotations starts with a successive interval. For example, Mode 3 starts with the third interval in Mode 1 (redrawn from Toussaint, 2013, Figure 14.8, with modifications).	15
2.5. The syncopation-shifted 3/2 metrical hierarchy applied to the standard pattern. The three lines inside the “clock” mark the strong beats. The arrows indicate how the weak beat preceding an absent strong beat takes the metrical weight of the strong beat.	22
2.6. Exchanging “halves” for the standard rhythms. The seven rhythms are arranged in the circular order of R4, R1, R5, R2, R6, R3, R7. The square brackets arrange the structure of each rotation into two types of approximate “halves” (5 + 7 and 7 + 5). The dotted lines mark the exchanges between “halves” to transform one rotation into a neighboring rotation.	24
2.7. Probe accent ratings of each rhythm (y-axis; full range [-100, 100]) in the standard pattern labelled by clock position (x-axis). Error bars indicate 95% confidence interval.	31
3.1. Exchanging “halves” for the diatonic modes. The seven modes are arranged in the circular order of M4, M1, M5, M2, M6, M3, M7. The square brackets arrange the structure of each rotation into two types of approximate “halves” (5 + 7 semitones and 7 + 5 semitones). The dotted lines mark the exchanges between “halves” to transform one rotation into a neighboring rotation.	42
3.2. Probe tone ratings of each mode (y-axis; full range [-100, 100]) in the diatonic pattern labelled by tone identity (x-axis), with ascending on the left and descending on the right. Error bars indicate 95% confidence	

interval.	49
4.1. Correlation between the fit judgment from Experiment 3 (y-axis) and the hierarchical-level match that is derived from the probe accent ratings in Experiment 1 and the probe tone ratings in Experiments 2 (x-axis). The fit lines indicate the correlation trends for individual participants.	60
5.1. Theoretical structure of the bell pattern in clock position. The black circles mark temporal events that sounded in the bell pattern. Starting from clock position 0 going clockwise, the lines connecting consecutive black circles form durations of 2 2 3 2 3 (redrawn from Toussaint, 2013, Figure 14.7).	66
5.2. Theoretical structure of the pentatonic pattern in tone identity. The black circles mark pitches that sounded in the pentatonic pattern. Starting from the pitch C going clockwise, the lines connecting consecutive black circles form semitone intervals of 2 2 3 2 3 (redrawn from Toussaint, 2013, Figure 14.7, with modifications).	67
5.3. The five unique rhythms of the bell pattern in clock position. Each of the five rotations starts with a successive duration. For example, Rhythm 2 starts with the second duration in Rhythm 1 (redrawn from Toussaint, 2013, Figure 14.7, with modifications).	68
5.4. The five unique modes of the pentatonic pattern in tone identity. Each of the five rotations starts with a successive interval. For example, Mode 3 starts with the third interval in Mode 1 (redrawn from Toussaint, 2013, Figure 14.7, with modifications).	69
5.5. Probe accent ratings of each rhythm (y-axis; full range [-100, 100]) in the	

bell pattern labelled by clock position (x-axis). Error bars indicate 95% confidence interval.	78
6.1. Probe tone ratings of each mode (y-axis; full range [-100, 100]) in the pentatonic pattern labelled by tone identity (x-axis), with ascending on the left and descending on the right. Error bars indicate 95% confidence interval.	86
7.1. Correlation between the fit judgment from Experiment 6 (y-axis) and the match of salience that is derived from the probe accent ratings in Experiment 4 and the probe tone ratings of the ascending and ascending followed by descending modes in Experiments 5 (x-axis). The fit lines indicate the correlation trends for individual participants.	92

LIST OF TABLES

2.1. Structure and stimuli of the standard and diatonic patterns.	19
2.2. Exchanging “halves” between the standard rhythms and between the diatonic modes.	25
2.3. The swapping process in the standard rhythms and diatonic modes.	26
2.4. Standard rhythms and diatonic modes arranged according to swap distance.	27
2.5. Coding and results from the mixed-model linear regression analysis on features of the standard pattern.	33
3.1. Coding and results from the mixed-model linear regression analysis on basic features of the diatonic pattern.	40
3.2. Species of the 4th and species of the 5th in the diatonic pattern: a) Types of Species of the 4th, b) Types of Species of the 5th, and c) Possible combinations for each mode.	44
5.1. Structure and stimuli of the bell and pentatonic patterns.	72
5.2. The swapping process for the bell rhythms and pentatonic modes.	74
5.3. Bell rhythms and pentatonic modes arranged according to swap distance.	74
8.1. Perception of the bell rhythms as subsets of the standard rhythms, and the pentatonic scale as subsets of the diatonic scales.	98

CHAPTER 1

INTRODUCTION

Music is a shared aesthetic experience of humans across cultures and throughout time. People engage with music for various reasons, such as to dance, to lighten the mood, to bond with family and friends, to perform a religious ritual, to focus attention, to retrieve a memory, to appreciate the composition, and so on. Because music-related activities are ubiquitous in daily life, music is a versatile access point to inform us about human perception and cognition.

The perception and cognition of music is of great theoretical and experimental interest in psychology and cognitive science for a number of reasons. First, it contributes to our understanding of the innerworkings of cognition, such as attention, memory, pattern abstraction, learning, emotion, and social mechanisms. Second, it provides rich, structured, and ecological stimuli to a wide variety of psychology experiments, and innovations in experimental methods benefit related research areas, such as psycholinguistic and event segmentation, just to name a few. Third, it provides empirical evidence for a more comprehensive understanding of neuroscience, cross-species comparison, nature and nurture, and well-being. Fourth, it contributes to interdisciplinary research by connecting cognition to disciplines such as music theory (the study of the practices and possibilities of music), acoustics (the science of sound), and kinematics (the geometry of motion). Fifth, it helps us better understand human's aesthetic preferences, which informs composition, musical performance, and audiovisual integration.

This dissertation focuses on the perception and cognition of the two most fundamental dimensions in music, pitch and time. Both dimensions have been studied intensively. It is important to note that a crucial part of our understanding is rooted in theoretical concepts. These theories provide important resource for empirical research because they inform precise testable

hypotheses that are ideal for quantitative testing. In preparation for the experiments presented in Chapters 2-7, this chapter will introduce the basic components of pitch (the vertical span of tones), time (the horizontal layout of tones), and the perception of each dimension. I will also introduce the topic of how pitch and time combine perceptually.

Rhythm and Metrical Hierarchy

When listening to music, we hear the serial layout of temporal events, which is the *rhythm*. A rhythm consists of a series of events marked by sound onsets. The time interval between one onset and the following onset is the *tone duration*. Tone durations often form simple ratios, such as 1:1 and 2:1. Thus, some temporal events are relatively long and others are relatively short. The ordering of these tone durations is specific for a rhythm. Listeners tend to hear simple ratios even when the duration ratio is not simple, suggesting categorical perception for rhythm (Clarke, 1987). In addition, they had difficulties distinguishing durations that do not cross the categorical boundary.

With exposure to a variety of rhythms, listeners perceive the structures underlying rhythms, called the *metrical hierarchies*. A metrical hierarchy consists of alternating strong and weak beats. These beats are regular pulses that we feel when listening to a rhythm. Strong beats feel more stressed than weak beats, and weak beats are heard in reference to strong beats. Each metrical hierarchy can have two or more levels of stress. Lerdahl and Jackendoff (1983) created a visual representation of metrical hierarchies, where strong beats are marked with more dots stacked at one temporal position and weak beats are marked with fewer dots. Figure 1.1 shows the metrical hierarchy of a phrase in *Let it be* by the Beatles, an example used by Temperley (1999). Temporal positions with four dots are the strongest beats, corresponding to syllables *find* and *Moth-*. Positions marked with three dots are also strong, followed by weaker positions with

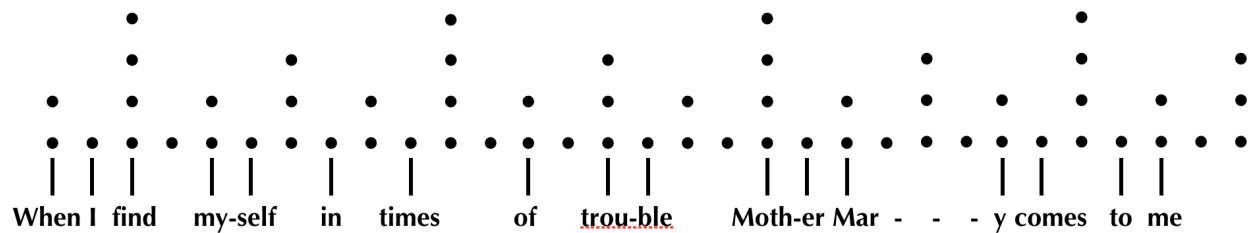


Figure 1.1. Metrical hierarchy (Lerdahl & Jackendoff, 1983; Temperley, 1999 with modifications).

two dots and the weakest positions with one dot.

A number of differences distinguish a rhythm from a metrical hierarchy. A rhythm exists in the form of actual sound. It is the surface form in which temporal events are arranged. Any change in tone duration alters the rhythm. A rhythm is also transient in that repetition is not a built-in characteristic. In contrast, a metrical hierarchy is a perception, not a collection of sound. It is the underlying structure of stresses that remains stable over time. Changes in the rhythm or the absence of a rhythm do not usually alter the metrical hierarchy. A metrical hierarchy is also cyclic, where it repeats from the beginning once it reaches the end.

A metrical hierarchy is initially generated based on the serial arrangement of long and short tones in a rhythm, which is a bottom-up process. However, with exposure to various musical rhythms, we form a collection of metrical hierarchies. When listening to a new rhythm, we tend to find a suitable metrical hierarchy and apply it to the rhythm, which is a top-down process. London (2004) called these two processes period extraction and template matching, respectively. Generally speaking, accented events in the rhythms tend to occur at the strong beats in metrical positions. However, the mapping between rhythm and metrical hierarchy is not one-to-one.

Different metrical hierarchies can be applied to the same rhythm, and various rhythms can be generated based on the same metrical hierarchy.

Rhythm and metrical hierarchy sometimes conflict with each other, creating a *syncopation*. A syncopation happens when a temporal event takes place at a weak beat whereas no temporal event takes place at the strong beat that precedes or follows the weak beat. For example, in Figure 1.1, *Mar-* takes place at a weak beat with two dots, and no syllable takes place at the strong beat with three dots following the weak beat. This local conflict forms a syncopation, and the temporal position of *Mar-* is the syncopated beat. A syncopation feels off-balanced compared to a piece where temporal events take place at strong beats, such as marching band music.

Finally, both rhythm and metrical hierarchy have been described in terms of *well-formedness* principles. These are theoretical rules that characterize how good a pattern is. They were proposed in Lerdahl and Jackendoff's (1983) *Generative Theory of Tonal Music*, where specific rules are summarized about the perceived structure that underlies the musical surface. The rules concern the relationship between temporal events, beats, and various levels of metrical stress. London (2004) further proposed a set of metric well-formedness constraints that define the widest possible range of metrical hierarchies across different musical traditions. In addition, Toussaint (2013) proposed geometrical characteristics that make a rhythm a *good* rhythm. These characteristics suggest that structures form good rhythms when each temporal event is mathematically unique and the entire pattern is balanced.

Scale and Tonal Hierarchy

The tones we hear in a piece of music are sounded at various frequencies, which are called *pitch*s. Pitch is considered to have two dimensions: chroma and height (Shepard, 1982). Figure 1.2 shows the chroma circle, which consists of twelve tones, forming an octave. The distance

between every two adjacent tones is a *semitone*, which corresponds to the relationship between any two adjacent keys on the piano keyboard. Tones with the same chroma can have different heights. For example, the middle C (C4) is one octave lower than C5 and has half of the frequency of C5.

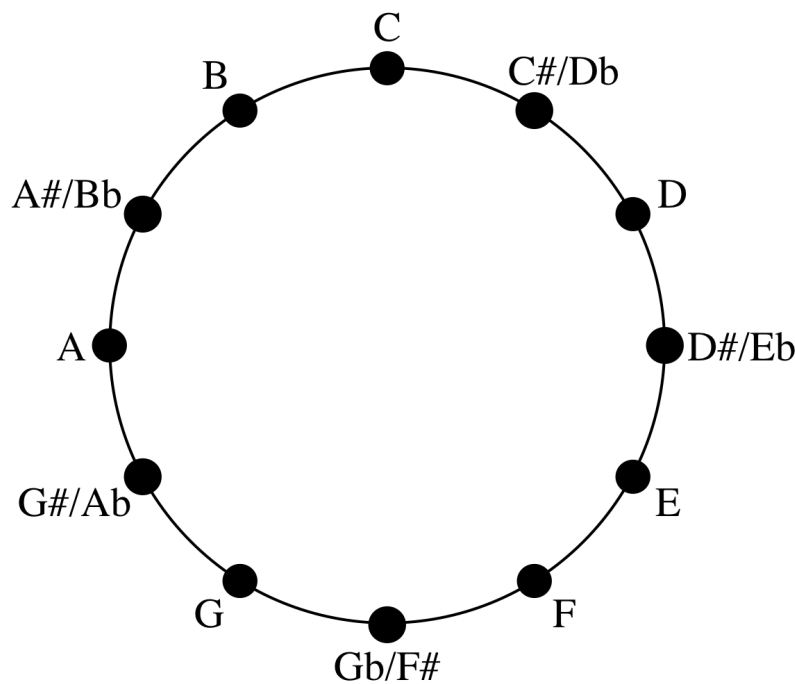


Figure 1.2. The chroma circle.

Most pitches that occur in a piece of music are drawn from a subset of chromas and can be arranged into a *scale*. A scale is a collection of pitches starting and ending on a certain chroma, spanning an octave, usually arranged from low to high frequencies (Power, 2001). The distance between adjacent pitches is called *interval size*. Interval sizes often form simple ratios, such as 2:1 and 3:1. Thus, some intervals are relatively large and others are relatively small. The ordering of these interval sizes is specific for a scale. Interval sizes affect listeners perception of music, such as grouping (Lerdahl & Jackendoff, 1983) and melodic expectancy (Narmour, 1990).

Pitch perception has been extensively studied. Collections of pitches form musical *keys*, a perceived hierarchical structure surrounding one pitch (chroma). The different keys are perceptually related to each other. Figure 1.3 shows the circle of fifths, where all major and minor keys are included. Keys that are close to each other are perceived as sharing more similarities than those that are far from each other.

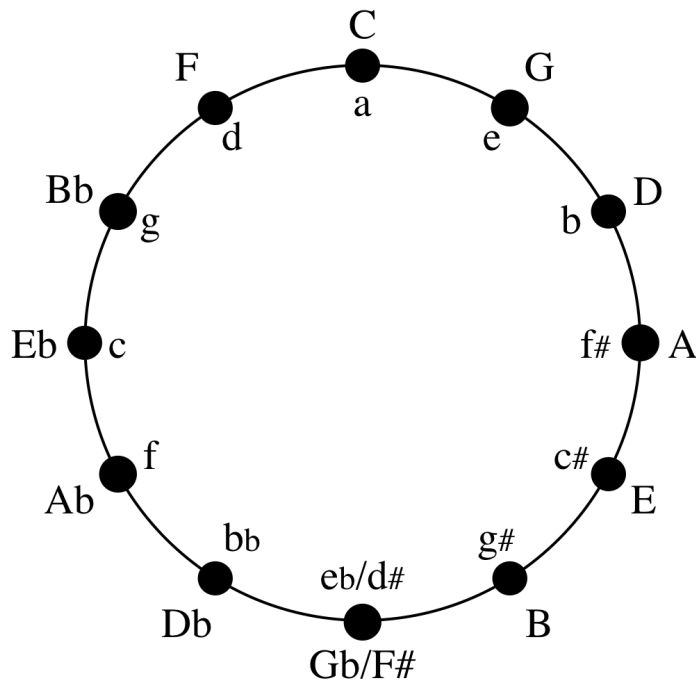


Figure 1.3. Circle of fifths with major keys in upper case letters and minor keys in lower case letters.

The one pitch that a key is named after is the *tonic*. The tonic is considered the cognitive reference point or the prototype of the key (Rosch, 1975; Rosch & Mervis, 1975; Krumhansl, 1990). For example, C is the tonic of the key C major. All other pitches in C major are more like the tonic, C, than C is like any other pitch. Thus, all other pitches in a key are perceived in reference to the tonic, forming a *tonal hierarchy*. Pitches that form consonant intervals (e.g.,

perfect fifth, perfect fourth, major and minor triads) with the tonic are more prototypical in the tonal hierarchy, and those that form dissonant intervals (e.g., tritone, minor second) with the tonic are less prototypical.

An experimental method was developed to study tonal hierarchy, called *the probe tone technique* (Krumhansl & Shepard, 1979). A sequence of pitches is played that imply a clear key. Following this tonal context and a short silence, a single pitch is played, called the *probe tone*. Listeners were asked to rate to what extent the probe tone fits into the preceding context. Higher ratings indicate stability of that pitch, that is, how prototypical a pitch is in the preceding key. Figure 1.4 shows the probe tone profile of the C major and C minor keys (Krumhansl & Kessler, 1982). The tonic was perceived as the most stable tone in a key, followed by pitches that form consonant intervals with the tonic (i.e., G and E for C major, Eb and G for C minor in the probe tone ratings; G and F for C major in consonance measures). Pitches that form dissonant intervals are perceived as the least stable tones (e.g., F#, C#).

People form tonal hierarchies even for unfamiliar musical styles, as evidenced in a number of studies using the probe tone technique (Krumhansl, 2000). For example, style-appropriate tonal hierarchies were found for atonal music (Krumhansl, Sandell, & Sergeant, 1987), Balinese music (Kessler, Hansen, & Shepard, 1984), North Indian music (Castellano, Bharucha, & Krumhansl, 1984), and specially composed music with a mixture of familiar and unfamiliar tonal structures (Oram & Cuddy, 1995). Regardless of musical training and cultural background, listeners perceptually organized pitches in a musical context based on tone distributions.

In addition, pitch relations have been studied based on Gestalt principles. Bregman (1990) developed Auditory Scene Analysis where principles in visual perception are applied to auditory perception. Listeners found tones that are in *proximity* to each other in pitch height to form

coherent melodies, but not pitches scrambled across octaves (Deutsch, 1995). Based on the preference for *similarity*, listeners group together sounds of music instruments that belong to the same section, such as the strings, the woodwinds, etc. Phrases that end on the tonic give listeners a sense of *closure*, which is preferred over phrases that do not end on the tonic. Series of pitches that form small incremental interval sizes in one direction establishes a *good continuation*. This is preferred to abrupt changes of directions between pitches with large interval gaps.

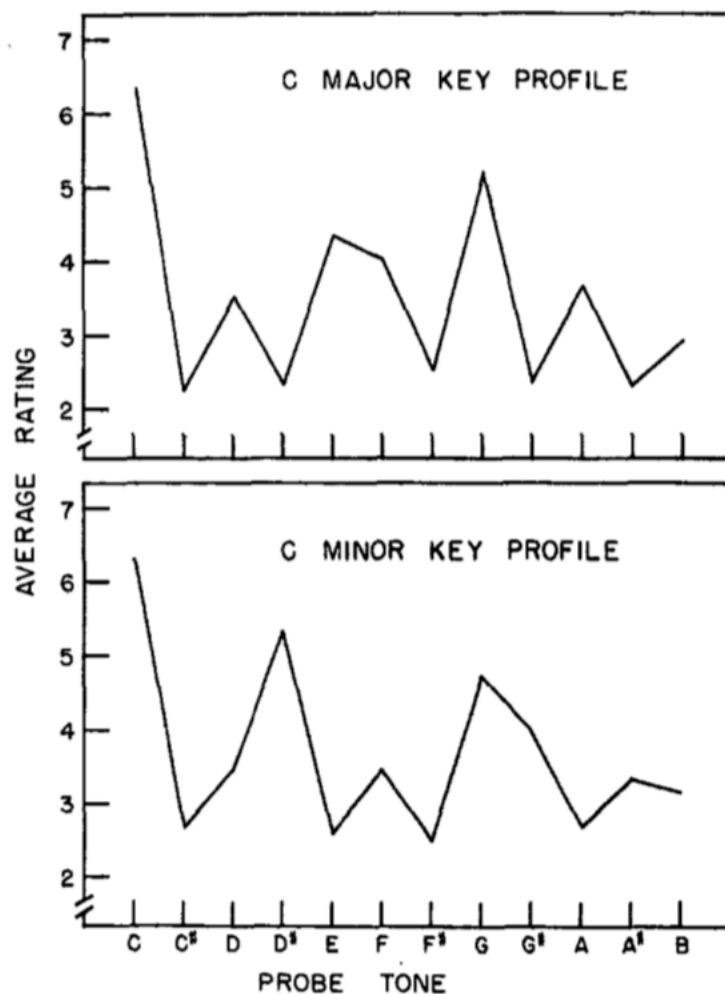


Figure 1.4. Probe tone profile for C major and C minor keys (Krumhansl & Kessler, 1982).

Pitch-Time Interaction

The relationship between dimensions has been studied in terms of separable versus interactive dimensions. Garner's (1976) sorting task is an example in vision. Participants were able to sort blue from red without being influenced by the shape of the images, showing that hue and shape are two separable dimensions. However, they could not sort blue from red without being influenced by the saturation of the images, suggesting that hue and saturation are two interactive dimensions.

The perceptual relationship between dimensions in music can be studied under a similar framework. Pitch and time are the two most structured dimensions in music. They have often been studied separately. However, because of the different terminologies, concepts, and structures in the two dimensions, pitch-time relationship has not been as extensively studied compared to the perception of each dimension. Depending on the task and stimulus materials, evidence supported the separable view or the interactive view.

A number of studies supported the separable view. For example, pitch and time are argued to be processed using different neural resources (Peretz & Coltheart, 2003; Peretz & Morais, 1989). However, the manipulation of individual tones and durations and tasks such as change detection (e.g., Lelo-de-Larrea-Mancera, Rodríguez-Agudelo, & Solís-Vivanco, 2017) tend to focus listeners' attention on isolated, local, and surface features, where separable processes are sufficient. Detailed review will be provided in Chapter 4.

On the other hand, evidence supporting the interactive view has demonstrated the effect of change in one dimension on the other dimension. For example, manipulating the predictability of the temporal context affected pitch judgment (e.g., Jones, 1993; Jones & Boltz, 1989; Jones, Boltz, & Kidd, 1982), and changes in pitch factors (e.g., contour, melody, harmony) also

influenced judgments of meter (Hannon, Snyder, Eerola, & Krumhansl, 2004). Another way of demonstrating pitch-time interaction is by measuring the preference for the co-occurrence of pitch and time features. For example, pitches that occurred more frequently at strong rather than weak metrical positions received higher ratings on tonal stability (Rosenthal & Hannon, 2016). Overall, the tasks often focus listeners' attention on integrated, global, and abstract features, which require interactive processes instead of separable ones. Further review to be detailed in Chapter 4.

CHAPTER 2

ISOMORPHISM IN THE STANDARD AND DIATONIC PATTERNS

The series of three experiments reported in Chapters 2, 3, and 4 was initially motivated by the question of how the two most important dimensions of music, pitch and time, interact in perception. To explore this, we built on a theoretical isomorphism between a rhythmic pattern and a scale pattern that was noted by Pressing (1983) and has received extensive theoretical discussion. The rhythmic pattern is often called *the standard pattern* and is the most pervasive rhythm in sub-Saharan Africa; the scale pattern is *the diatonic pattern* and is the most prominent scale in Western tonal music. Whereas in Western music, the progression of tone, such as harmony and melody, is considered primary and rhythm is considered secondary, Chernoff (1979) noted that the reverse sensibility exists in African music, where rhythm is considered primary.

The two most prominent patterns in their respective cultures, the standard pattern and the diatonic pattern, are isomorphic based on their shared surface structure of 2 2 1 2 2 2 1. This structure describes both the temporal intervals between every two consecutive events in the standard pattern (Figure 2.1), and the pitch intervals between every two consecutive tones in the diatonic pattern (Figure 2.2). The rhythmic and scale patterns share a number of interesting theoretical properties, to be detailed later, but it should be noted here that when the pattern begins with each successive duration, seven unique patterns are generated. For example, Rhythm 1 has the structure of 2 2 1 2 2 2 1, Rhythm 2 starts with the second duration in Rhythm 1, generating the rotation 2 1 2 2 2 1 2, and so on (see Figure 2.3). Pressing (1983, Table I) and Agawu (2006) report that all seven rotations of rhythms are found in African music. Similarly, the seven rotations of the scale, sometimes referred to as “church modes” in the context of

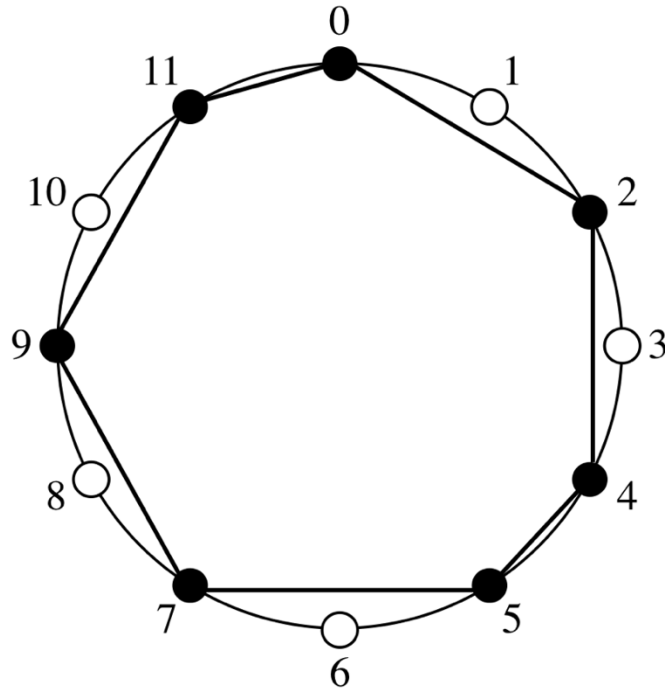


Figure 2.1. Theoretical structure of the standard pattern in clock position. The black circles mark temporal events that sounded in the standard pattern. Starting from clock position 0 going clockwise, the lines connecting consecutive black circles form durations of 2 2 1 2 2 2 1 (redrawn from Toussaint, 2013, Figure 14.8).

Western music, are also all unique (see Figure 2.4). Pressing (1983) suggested that the isomorphism between the rhythms and the modes reflect common underlying cognitive principles. However, London (2002) questioned the significance of the isomorphism based on differences in geometrical models of pitch relationships and temporal relations.

Three experiments were conducted to find out to what extent the standard rhythms and the diatonic modes are perceptually isomorphic. We began by asking whether listeners would prefer combinations of modes and rhythms in which larger pitch intervals are played with longer tone durations (inter-onset intervals). In other words, would the seven isomorphic combinations

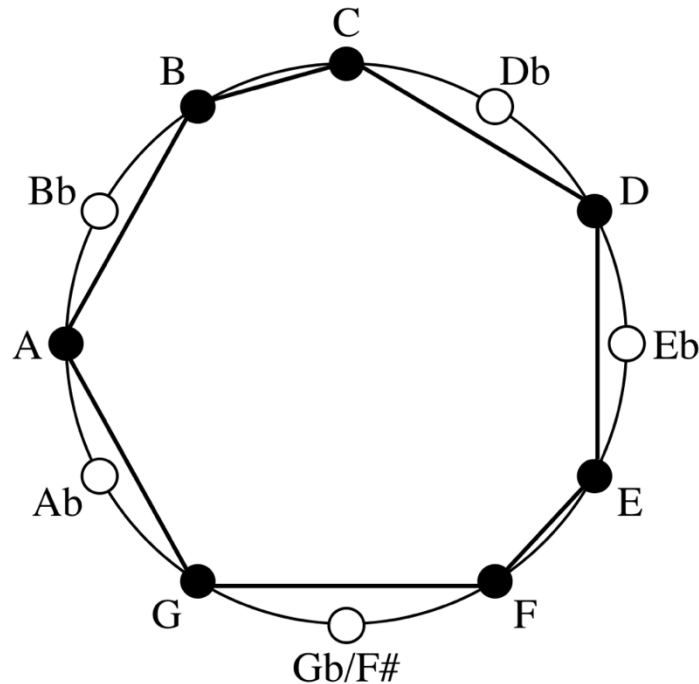


Figure 2.2. Theoretical structure of the diatonic pattern in tone identity. The black circles mark pitches that sounded in the diatonic pattern. Starting from the pitch C going clockwise, the lines connecting consecutive black circles form semitone intervals of 2 2 1 2 2 2 1 (redrawn from Toussaint, 2013, Figure 14.8, with modifications).

be preferred over other combinations in which the surface structures are not isomorphic?

Alternatively, preferences for rhythm-mode combinations may depend on how the hierarchical organization of the modes and rhythms are aligned relative to one another. Would they prefer combinations in which tonally strong events occur at points of rhythmic stress?

Examining this latter possibility required first empirically determining the hierarchies of events in the rhythms and the hierarchies of tones in the modes. Various music theoretical proposals were compared with the perceived hierarchical organizations. Also tested were

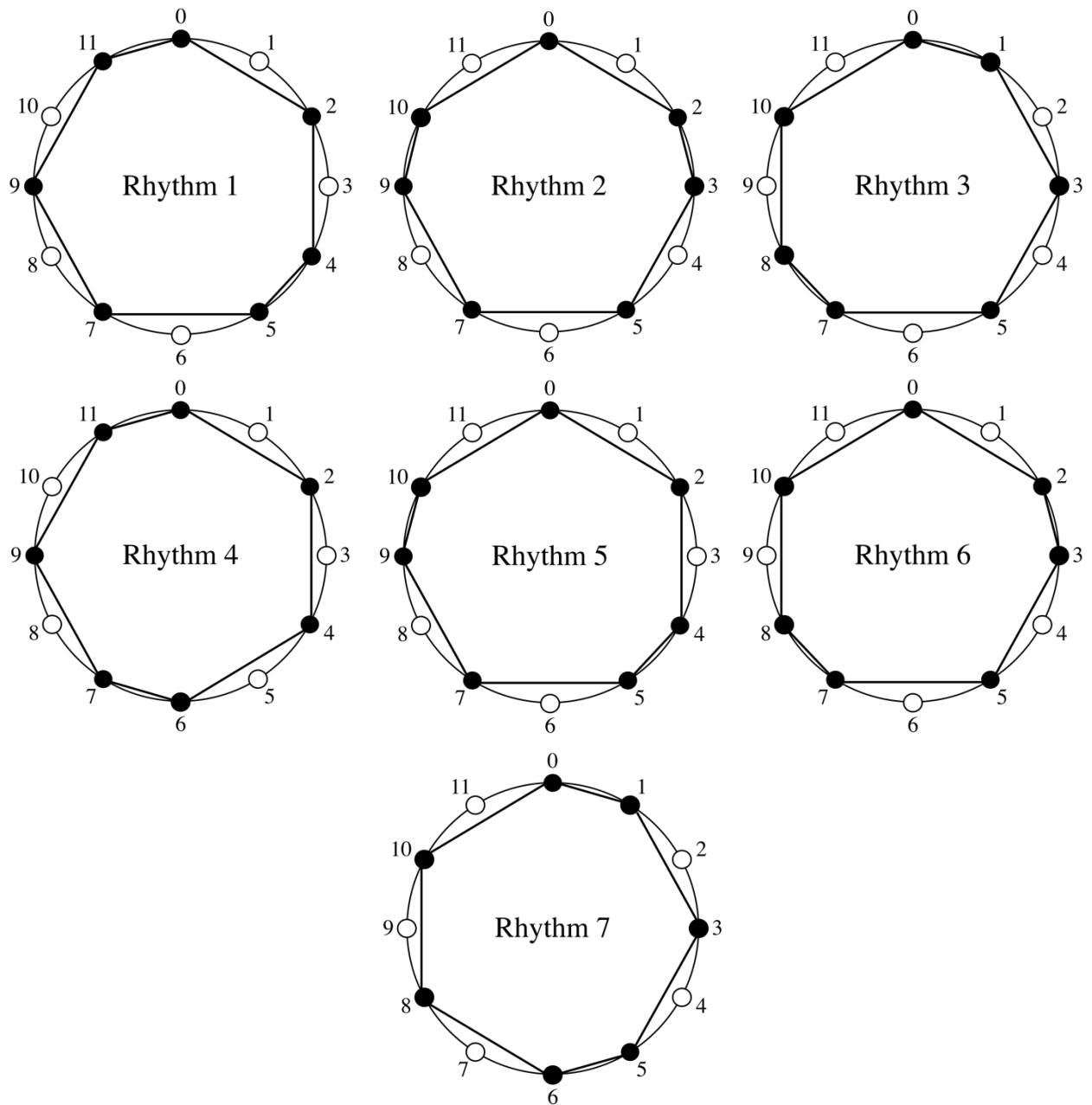


Figure 2.3. The seven unique rhythms of the standard pattern in clock position. Each of the seven rotations starts with a successive duration. For example, Rhythm 2 starts with the second duration in Rhythm 1 (redrawn from Toussaint, 2013, Figure 14.8, with modifications).

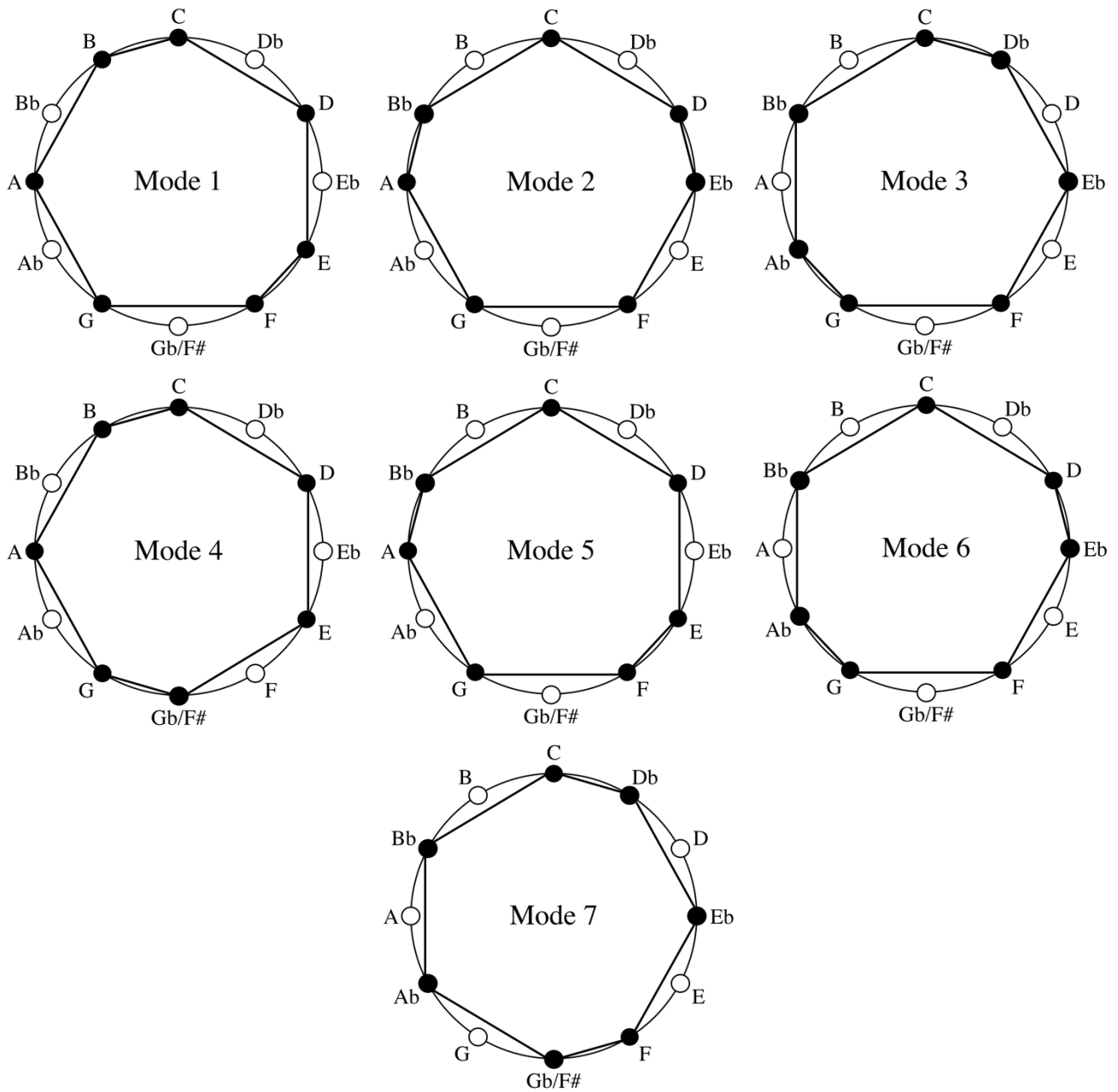


Figure 2.4. The seven unique modes of the diatonic pattern in tone identity. Each of the seven rotations starts with a successive interval. For example, Mode 3 starts with the third interval in Mode 1 (redrawn from Toussaint, 2013, Figure 14.8, with modifications).

predictions about the relationships between the rhythms and between the modes.

A method needed to be devised to probe the hierarchical structure of the standard rhythms; this was done by analogy to tonal hierarchies. In tonal music, one or more pitches are described as stable reference points, with other tones heard in relation to them. The tonic is generally considered the most important or stable tone in major and minor scales; other tones are heard as more or less related to it, creating a hierarchy (Krumhansl, 1990). By analogy, certain time points in rhythms may be heard as stable reference points, with other time points heard in relation to them. For example, an up-beat in Western music is heard in relation to the beat that follows it. To obtain a perceptual measure of the stability of each event in the rhythms, we dynamically accented one of the tones in the rhythms (played monotonically) and asked how well that accent fits into the rhythm. This was done for all seven temporal positions in each of the seven rhythms. Experiment 1 is the probe accent experiment using the method just described, which will be detailed shortly in this chapter.

Experiment 2 uses the probe tone method in which a musical context, in this case, each of the church modes, is followed by each of the tones in the scale, and listeners rate how well that tone fits into the scale. Although the technique has been applied extensively to major and minor contexts, to our knowledge it has not been applied to all seven of the church modes; this is what is done in Experiment 2, which will be detailed in Chapter 3.

Finally, in Experiment 3, listeners judged how well the rhythm fits the scale for all possible combinations of rhythms and modes. The analyses considered the influence of surface isomorphisms between tone duration and interval size. That is, are combinations preferred in which larger pitch intervals are separated by longer durations? Alternatively, are preferences driven by matches between rhythmic and tonal hierarchies? This would be the case if tonally

stable tones (as measured in Experiment 2) occur at rhythmically stable points in time (as measured in Experiment 1). Together with Experiment 3, Chapter 4 will discuss the results of Experiments 1, 2, and 3. The discussion includes differences and similarities in the cognitive principles underlying the perception of pitch and time, and future directions.

EXPERIMENT 1

As noted above, the standard pattern, pervasive in African music, can be viewed as a rhythmic sequence with the temporal intervals of 2 2 1 2 2 2 1. Originating in Southern Ewe dance music in Ghana, the standard pattern is found south of the Sahara throughout West and Central Africa, specifically, geographical regions that are demarcated by specific Niger-Congo languages (Agawu, 2006; Kubik, 1999; Pressing, 1983; Rahn, 1987; Temperley, 2000; Toussaint, 2013).

The standard pattern, and related rhythms, are played in a repetitive and continuous fashion throughout a piece and are often referred to as a form of rhythmic ostinato, or more specifically, a timeline (Agawu, 2006; Nketia, 1974; Rahn, 1987). Rhythmic ostinato refers to any continuously repeating rhythmic pattern, whereas timelines are more particular in that they are prominent in the music, are easily recognized and remembered, and function as a reference for musicians playing other rhythms simultaneously (Toussaint, 2013). Similar to Western meters, timelines have been viewed as a temporal reference (Agawu, 2006), “rhythms of reference” (Rahn, 1987), and “controlling structural concept” (Anku, 2000). The bell, which is the drum that the Ewes use to play the standard pattern, has been referred to as the “heartbeat” (Chernoff, 1979). However, unlike Western meters, where beats tend to be organized in evenly spaced groups (typically of 2’s or 3’s), timelines usually have unevenly spread temporal events, for

example, as described above for the standard pattern. This results in asymmetrical sub-patterns, which will be explained shortly.

Any of the seven events can serve as the first beat of the rhythm (Agawu, 2006; Pressing, 1983), and the starting point varies across cultures and geographical regions. Pressing (1983, Table I) summarized regions in West Africa where each of the seven rhythms occur. The standard pattern is *rotational*, such that the seven starting points generate seven unique patterns. The first three columns in Table 2.1 show the structure and the rhythmic notation of the seven rhythms, where Rhythm 1, Rhythm 2, ..., Rhythm 7 are abbreviated as R1, R2, ..., R7. Figure 2.3 illustrates the same rhythms in terms of clock position. However, it might be noted that, because the rhythm is played in a repetitive fashion during a piece, the starting point may not be very salient in a musical context.

Rhythms in the standard pattern can be characterized by three properties (Kubik, 1999): their *cycle number*, which is the number of possible temporal positions contained in the rhythm (i.e., 12 for the standard pattern), the *number of events* in the cycle, which is the number of temporal positions where an event is sounded (i.e., 7), and the *asymmetric distribution* generating two adjacent sub-patterns with unequal durations (i.e., for the 2 2 1 2 2 2 1 rotation, the first sub-pattern has a duration of 5, two 2's and one 1, and the second has a duration of 7, three 2's and one 1).

We are aware of no empirical research on how these rhythms are perceptually organized, either in terms of grouping or hierarchical patterns of stress. However, various authors have made proposals about how events in the standard pattern are perceived. One focused on tone

Table 2.1. *Structure and stimuli of the standard and diatonic patterns.*

Standard pattern			Diatonic pattern		
Structure	Rhythm	Temporal arrangement	Structure	Mode	Pitch arrangement
2 2 1 2 2 2 1	Rhythm 1 (R1)	♪♪♪♪♪♪♪ (J)	T T S T T T S	Mode 1 (M1)	C D E F G A B (C)
2 1 2 2 2 1 2	Rhythm 2 (R2)	♪♪♪♪♪♪♪ (J)	T S T T T S T	Mode 2 (M2)	C D E ^b F G A B ^b (C)
1 2 2 2 1 2 2	Rhythm 3 (R3)	♪♪♪♪♪♪♪ (J)	S T T T S T T	Mode 3 (M3)	C D ^b E ^b F G A ^b B ^b (C)
2 2 2 1 2 2 1	Rhythm 4 (R4)	♪♪♪♪♪♪♪ (J)	T T T S T T S	Mode 4 (M4)	C D E F [#] G A B (C)
2 2 1 2 2 1 2	Rhythm 5 (R5)	♪♪♪♪♪♪♪ (J)	T T S T T S T	Mode 5 (M5)	C D E F G A B ^b (C)
2 1 2 2 1 2 2	Rhythm 6 (R6)	♪♪♪♪♪♪♪ (J)	T S T T S T T	Mode 6 (M6)	C D E ^b F G A ^b B ^b (C)
1 2 2 1 2 2 2	Rhythm 7 (R7)	♪♪♪♪♪♪♪ (J)	S T T S T T T	Mode 7 (M7)	C D ^b E ^b F G ^b A ^b B ^b (C)

duration. Agawu (2006) suggested that the longer durations (i.e., 2) establish a sequence and the shorter durations (i.e., 1) interrupt and destabilize it. By constantly breaking and regaining the balance that has been set by the longer durations, the shorter durations introduce syncopation and move the rhythmic pattern forward. In this account, the standard pattern is dominated by runs of 2's which are broken up by individual 1's.

Another proposal focused on cycles. Pressing (1983) arranged the 12 temporal positions into categories by subdividing the standard pattern based on cycle of 2 (i.e., 1 2 1 2 1 2 1 2 1 2 1 2), cycle of 3 (i.e., 1 2 3 1 2 3 1 2 3 1 2 3), cycle of 4 (i.e., 1 2 3 4 1 2 3 4 1 2 3 4), and cycle of 6 (i.e., 1 2 3 4 5 6 1 2 3 4 5 6). For each cycle type, the events of the standard pattern (regardless of starting position) are roughly evenly distributed across the categories. According to Pressing, this makes the rhythmic pattern maximally versatile, so that a master drummer has many options for choosing a starting point. Agawu (2006) argued that cycle of 3 is culturally salient because it is what musicians, and especially dancers, feel and rely upon; it resembles how the dancers' feet move to the standard pattern. In contrast, Jones (1958) claimed that clapping duple and triple to the same song are both usual and natural to Africans, and the two groupings make no difference to the phrasing or accentuation of the pattern.

A third proposal about how events in the standard pattern are perceptually organized focuses on metrical hierarchies. A metrical hierarchy consists of nested levels that reflect the periodic pattern of relatively stronger and weaker beats (Lerdahl & Jackendoff, 1983; Palmer & Krumhansl, 1990). Derived from Pressing's subdivisions, Temperley (2000) proposed looking at the standard pattern with respect to three metrical hierarchies: 12/8 (i.e., 4 1 1 2 1 1 3 1 1 2 1 1), 6/4 (i.e., 4 1 2 1 2 1 3 1 2 1 2 1), and 3/2 (i.e., 4 1 2 1 3 1 2 1 3 1 2 1), where the larger numbers correspond to stronger beats and the smaller numbers correspond to weaker beats. Although

these metrical divisions are theoretically possible, most scholars have tended to consider the standard pattern, like many other Western African rhythms, to be in 12/8 metrical hierarchy (Pressing, 1983; Temperley, 2000). Agawu (2006) suggested the notation 12/8 is a convenient way to imply 4 cycles of 3.

A fourth proposal modifies the metrical hierarchies according to a principle called *syncopation shift* (Temperley, 2000). According to Temperley (1999), syncopation shift applies to the case in which no event takes place on a strong beat but an event takes place on the weak beat just before it. When this happens, the event that takes place on the weak beat takes on the metrical hierarchy weight of the following (empty) metrical position. In other words, the event is perceived as “belonging” to the strong beat. Temperley (2018) called this phenomenon *anticipatory syncopation* and noted its pervasiveness in not only rock but also jazz, blues, and Tin Pan Alley. Rahn (1987) suggested the phenomenon conflicts with the meter but Temperley proposed instead that it actually reinforces the meter.

Figure 2.5 illustrates how the 3/2 metrical hierarchy is modified by syncopation shift. The metrical hierarchy values for 3/2, starting from clock position 0 and going clockwise, are 4 1 2 1 3 1 2 1 3 1 2 1; they are shown in parentheses. The three solid lines inside the circles mark the strong beats in clock positions 0, 4, and 8. The arrows mark the syncopation shifts. For example, in Rhythm 1, because no event takes place at the strong clock position 8 (with the weight of 3), but an event takes place at clock position 7, it is assigned the metrical accent of clock position 8, changing its weight from 1 to 3. Thus, for this rhythm, the syncopation-shifted values are 4 2 3 1 3 1 1 for the seven events. Note that a syncopation shift occurs at least once in all seven rhythms.

Locke (1982) mentioned the importance of period (cycle) and accents within that cycle to create the sense of meter. In addition, the accent seems to be perceived as shifting frequently

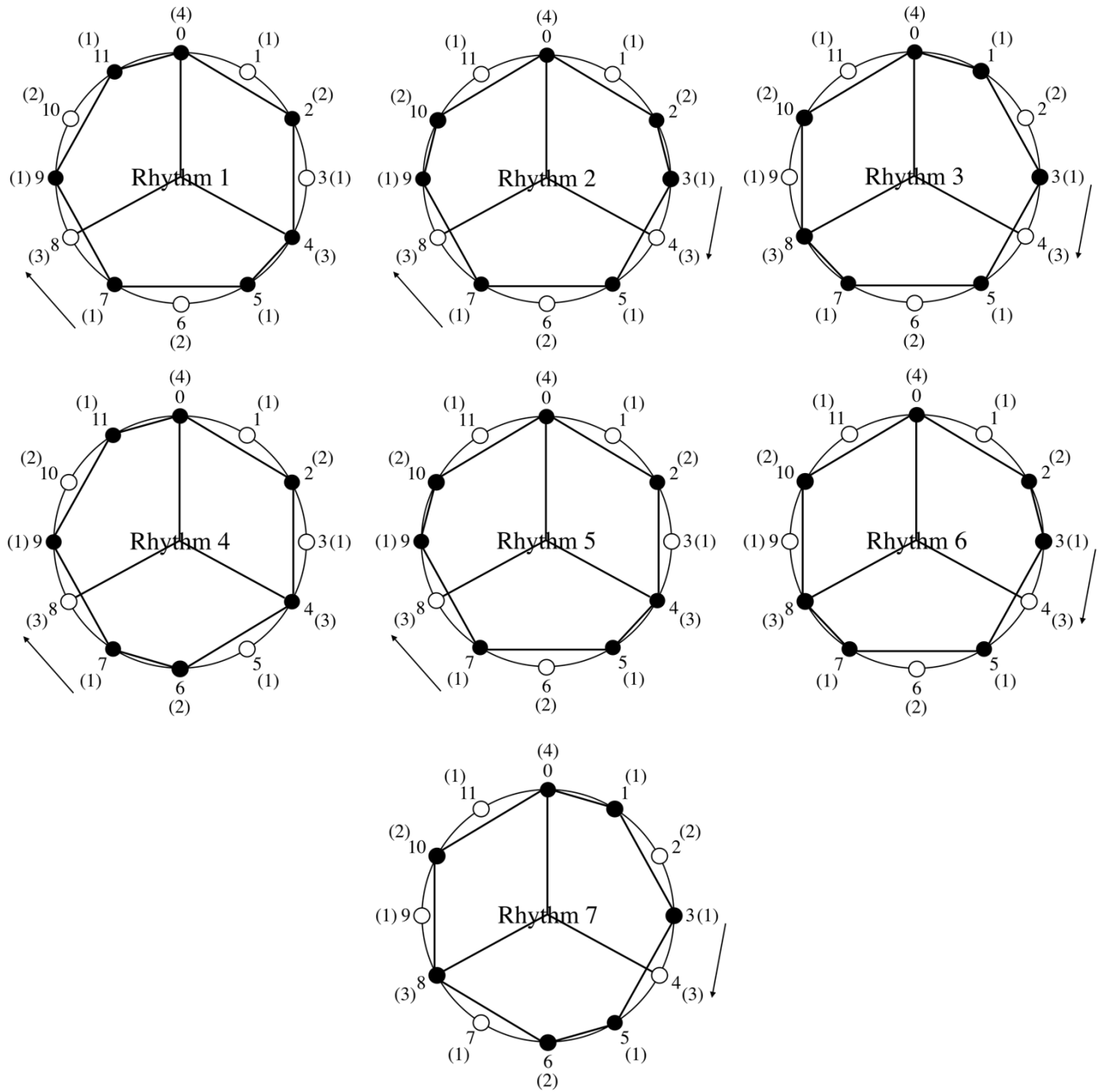


Figure 2.5. The syncopation-shifted 3/2 metrical hierarchy applied to the standard pattern. The three lines inside the “clock” mark the strong beats. The arrows indicate how the weak beat preceding an absent strong beat takes the metrical weight of the strong beat.

when a consistently accented offbeat is sufficiently strong. These principles would be supported if listeners prefer events at specific temporal positions across all rhythms, and if the location of the preferred accents shift based on the given rhythmical context. Based on Locke's (1982) ethnomusicological observations of African rhythms, Pressing (1983) suggested the possibility of multiple cognitive principles coexisting in the perception of the standard rhythms, which he called *multistability*.

In addition, there have been proposals about how the different rotations relate to each other. In other words, how might the distances between them be measured? A circular representation of how the rhythms are related to one another is derived by a process called *exchanging "halves"* (Agawu, 2006). Each rhythm can be viewed as two approximate "halves," or two sub-patterns as mentioned earlier. One half consists of two 2's and a 1, making a total duration of 5, and the other half consists of three 2's and a 1, making a total duration of 7 (see Table 2.1). By dividing each rhythm into 5 + 7 or 7 + 5 and then exchanging the two "halves," the seven rhythms can be arranged in the circular order R4, R1, R5, R2, R6, R3, and R7. Figure 2.6 shows this arrangement with the "halves" (5 and 7) marked in square brackets and the exchanges marked by dotted lines. For example, when R1 is divided into 5 + 7, exchanging the two "halves" resulted in R4; when R1 is divided into 7 + 5, exchanging the two "halves" resulted in R5. Table 2.2 shows distances between the seven rhythms in terms of the number of exchanges needed between "halves."

Swap distance is another possible measure of perceptual distances between rhythms (Toussaint, 2013). It is the minimum number of swaps between adjacent pulse units necessary to transform one rhythm into another. Swap distance only applies to rhythms with the same number of events. Table 2.3 shows the seven rhythms in the order of R4, R1, R5, R2, R6, R3, R7,

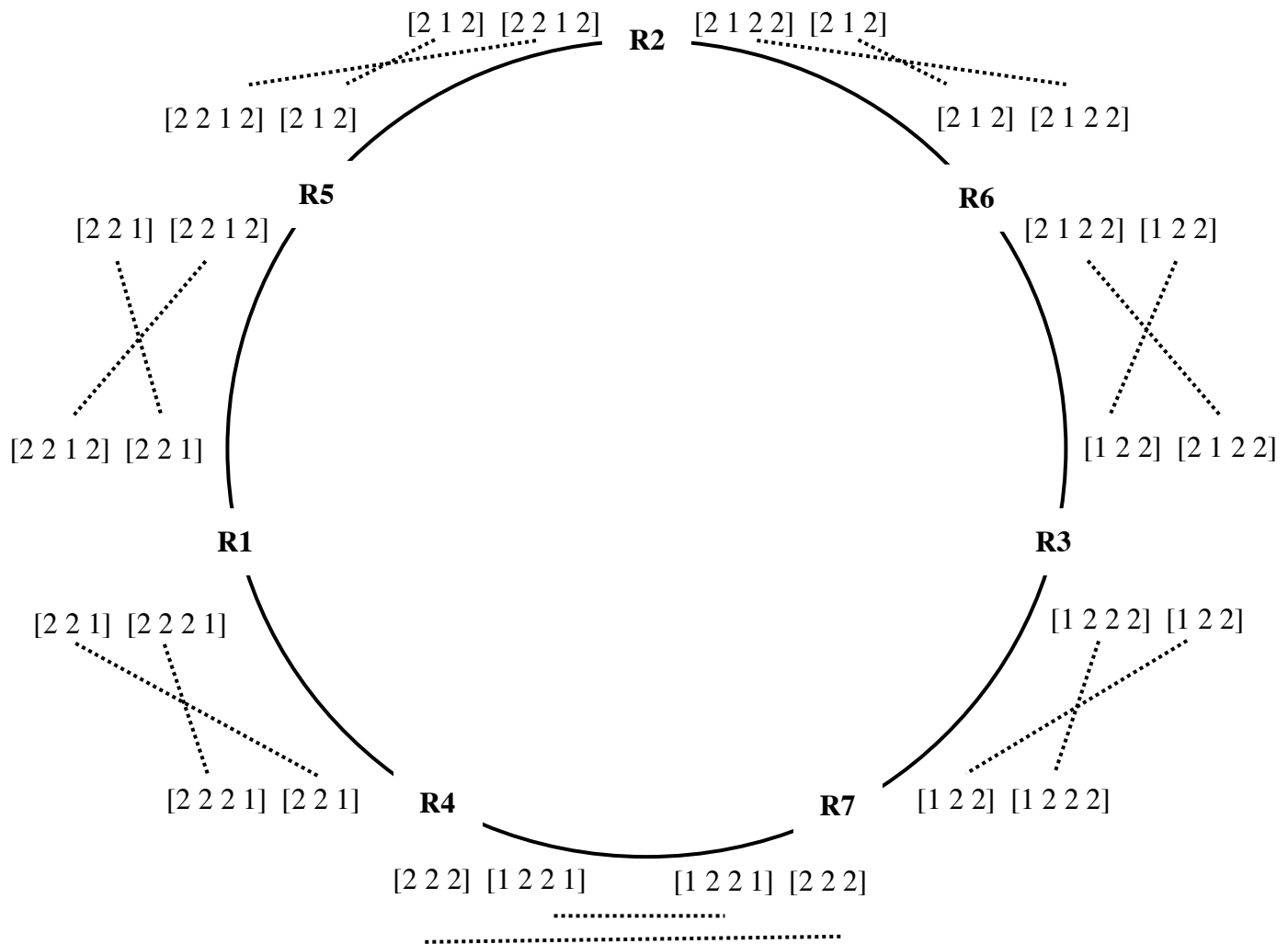


Figure 2.6. Exchanging “halves” for the standard rhythms. The seven rhythms are arranged in the circular order of R4, R1, R5, R2, R6, R3, R7. The square brackets arrange the structure of each rotation into two types of approximate “halves” (5 + 7 and 7 + 5). The dotted lines mark the exchanges between “halves” to transform one rotation into a neighboring rotation.

indicating each event by its clock position (1 = event, 0 = no event). Swapped positions are highlighted. For example, R4 can be transformed to R1 by swapping clock positions 5 and 6, making the swap distance 1. In turn, R1 can be transformed to R5 by swapping clock positions

Table 2.2. *Exchanging “halves” between the standard rhythms and between the diatonic modes.*

Rhythm	R4	R1	R5	R2	R6	R3	R7	
R4	0	1	2	3	3	2	1	M4(1#)
R1	1	0	1	2	3	3	2	M1(0#, 0b)
R5	2	1	0	1	2	3	3	M5(1b)
R2	3	2	1	0	1	2	3	M2(2b)
R6	3	3	2	1	0	1	2	M6(3b)
R3	2	3	3	2	1	0	1	M3(4b)
R7	1	2	3	3	2	1	0	M7(5b)
	M4(1#)	M1 (0#, 0b)	M5(1b)	M2(2b)	M6(3b)	M3(4b)	M7(5b)	Mode (no. of accidentals)

10 and 11, making the swap distance 1. Accordingly, the swap distance between R4 and R5 is 2 (1 + 1). In other words, R4 can be transformed to R5 by swapping both clock positions 5 and 6, and clock positions 10 and 11. Table 2.4 shows the swap distance between every two rhythms. These two distance measures order the rhythms the same except that exchanging “halves” is a circular representation and swap distance is a linear representation with R4 and R7 farthest apart.

Objectives of the Study

Experiment 1 was designed to explore the perceptual organization of the standard pattern, in particular, hierarchies of accents in the rhythms. Before describing the specifics of the method, it may be useful to discuss the difference between rhythm and meter. As reviewed by Vuust and Witek (2014), rhythm refers to the pattern of discrete durations and is considered to depend on processes of perceptual grouping. Time is organized sequentially and the perceptual organization depends on the pattern of temporal intervals. In contrast, meter contains a regular pulse or tactus,

Table 2.3. *The swapping process in the standard rhythms and diatonic modes.*

Standard rhythms in clock position													
Rhythm	0	1	2	3	4	5	6	7	8	9	10	11	
R4	1	0	1	0	1	0	1	1	0	1	0	1	M4(1#)
R1	1	0	1	0	1	1	0	1	0	1	0	1	M1(0#, 0b)
R5	1	0	1	0	1	1	0	1	0	1	1	0	M5(1b)
R2	1	0	1	1	0	1	0	1	0	1	1	0	M2(2b)
R6	1	0	1	1	0	1	0	1	1	0	1	0	M6(3b)
R3	1	1	0	1	0	1	0	1	1	0	1	0	M3(4b)
R7	1	1	0	1	0	1	1	0	1	0	1	0	M7(5b)
	C	Db	D	Eb	E	F	F#/Gb	G	Ab	A	Bb	B	Mode (no. of accidentals)
Diatonic modes in tone identity													

which is hierarchically differentiated into relatively strong and weak accents which recur with periodic regularity. Rhythmic events are perceived, at least in part, as accented according to the underlying metrical hierarchy. Rhythmic and metric accents often align but, according to London (2012), when rhythmic accent is not cued by duration or dynamics, the accents are often determined by the meter, a top-down process. This orientation seems to be generally accepted in research on metrical hierarchies, which is extensive (e.g., Ladinig, Honing, Haden, & Winkler, 2009; Large & Kolen, 1994).

However, an alternative process of bottom-up abstraction should also be considered. Listeners may impose a perceptual hierarchy on rhythms where some events naturally accrue perceptual accents independently of an underlying meter. In other words, meter might not be the

Table 2.4. *Standard rhythms and diatonic modes arranged according to swap distance.*

Rhythm	R4	R1	R5	R2	R6	R3	R7	
R4	0	1	2	3	4	5	6	M4(1#)
R1	1	0	1	2	3	4	5	M1(0#, 0b)
R5	2	1	0	1	2	3	4	M5(1b)
R2	3	2	1	0	1	2	3	M2(2b)
R6	4	3	2	1	0	1	2	M6(3b)
R3	5	4	3	2	1	0	1	M3(4b)
R7	6	5	4	3	2	1	0	M7(5b)
	M4(1#)	M1 (0#, 0b)	M5(1b)	M2(2b)	M6(3b)	M3(4b)	M7(5b)	Mode (no. of accidentals)

only factor governing rhythmic hierarchies. Although it is thought that perceptual grouping is important for perceiving rhythms, there is no fixed relationship between grouping and metrical accent (Cooper & Meyer, 1960; Krumhansl, 2000; Lerdahl & Jackendoff, 1983). The first and last elements of a group may be heard as accented, as might longer tones or tones that precede gaps, or the first of a series of similar elements (Garner, 1974). Rhythmic events may also be accented by other factors, such as pitch contour, harmonic changes, and melodic parallelism (Lerdahl & Jackendoff, 1983). Although theories exist regarding rhythmic accents independently of meter, there is relatively little empirical evidence.

In order to probe specific temporal positions in the standard pattern, we dynamically accented one of the events in each rhythm by playing it louder than the other events in the rhythm. This is called the *probe accent*. The listeners' task was to rate how well the probe accent fits into the rhythm. We did this for each of the first seven events in each rhythm. The

assumption behind this method was that if a structurally stable or relatively strong event is dynamically accented, listeners will judge that the accent fits well into the rhythm. That is, if an event would usually be heard as accented perceptually, then actually accenting it would seem natural. However, if an event that is usually perceived as weak, an accent on it would not be judged as fitting well. The resulting data will be considered in light of the various proposals summarized above for how events within the rhythms are organized and how they are related to one another.

Method

Participants

Forty-five students (7 males and 38 females, age 18-40) participated in Experiment 1 for course credit or \$5 cash. On average, they have played music for 12.79 years (including music lessons; summed over all instruments and voice; range, 1-31 years; one person had no musical training). The study was approved by Cornell University's Institutional Review Board.

Apparatus and Stimulus Materials

We constructed seven rhythms based on the standard pattern by shifting the starting tone duration (see Figure 2.3 and Table 2.1). All rhythms consisted of 8 monotonic tones on C4, forming 7 time intervals. On each trial, only one temporal position was accented by increasing its loudness by 6.5 decibels, forming the probe accent. The probe accent was on one of the first seven events in each rhythm. All rhythms were at the tempo of 140 quarter-note per minute (each quarter note lasts approximately 440 ms) and were sounded in piano timbre. On each trial, the rhythm was played twice separated by a silence that lasted 3 quarter notes (each trial lasted approximately 7.48 s). This repetition allowed the participants to hear where the accent is on the first hearing, so they can focus on how it fits into the rhythm on the second hearing. Also

constructed were 7 unaccented rhythms, called the *neutral* rhythms, which were identical to the accented rhythms, except that all the tones were played at the same loudness level.

Procedure

Participants rated how well the probe accent fits into the rhythm by moving a slider on a continuous scale ranging from extremely bad fit (left end) to extremely good fit (right end). They began with four practice trials. Next, they listened to each of the 49 probe accent trials, which were blocked by rhythm type (i.e., R1, R2, etc.). Each block began with the neutral rhythm in order to familiarize the listener with the rhythm in the block. The 7 blocks were presented in a randomized order within which the 7 probe accent trials were played in randomized order. Following all the probe accent trials, participants listened to each of the 7 neutral (unaccented) rhythms in a randomized order and rated them on two scales: 1) how familiar the rhythm was prior to the experiment, and 2) how well-formed the rhythm seems, in other words, whether or not the rhythm forms a good pattern. At the end of the study, they filled out a demographics questionnaire. The experiment lasted approximately 30 minutes.

Results

All continuous rating scales were coded from -100 to 100. Ratings from only 49% of participants correlated significantly with the average probe accent ratings, suggesting there may be some individual differences in the pattern of responses. To investigate this, a clustering analysis (Ward method) was done on the individual data and three main clusters were found. These were compared to demographic factors, including years of training on musical instruments or voice, whether the training included piano, whether they have taken a music theory class, whether they have been in a class where African rhythms were mentioned, whether they listen to classical music or not, and their familiarity with music or dance related to the standard pattern

(Afro-Caribbean music, Jazz, Latin Dance, and African Dance). However, only two of these were different between the groups and the differences were only weakly significant (years of training: $F(2, 41) = 4.29, p = .02$; whether the training was on piano: $F(2, 41) = 3.60, p = .04$).

A related question is whether some rhythms were more familiar than others prior to the experiment. To test this, we conducted a preliminary analysis of the familiarity ratings on the seven neutral (unaccented) rhythms and found significant differences between rhythms ($F(6, 264) = 4.98, p < .0001$). A post-hoc test (Tukey HSD) showed that R4 was rated as the most familiar and R7 was rated as the least familiar; the remaining rhythms were not statistically different from each other. As can be seen in Table 2.1, R4 is the rhythm that starts with the longest run of long durations. In contrast, R7 starts with a short duration followed by only two long durations before the next short duration. The same ordering was found for the well-formedness ratings which correlated significantly with familiarity ($F(1, 269) = 111.18, p < .0001$).

Hierarchies within the rhythms

Figure 2.7 shows the probe accent ratings for each of the seven rhythms plotted according to serial position and labeled by clock position so that durations can be determined by the values on the x-axis. For all the rhythms, there are peaks at the first and third serial positions. There is a third peak for all the rhythms except R7, which was judged the least familiar and well-formed. For R1, R2, R4, and R5, the third peak is at serial position 5. For R3 and R6, the peak is at serial position 6. These two are the only rhythms in which the event in serial position 6 is preceded by a short event in serial position 5, suggesting the accent is preferred on the long event following the short event.

The general pattern in the probe accent judgments of three peaks across the rhythms suggests

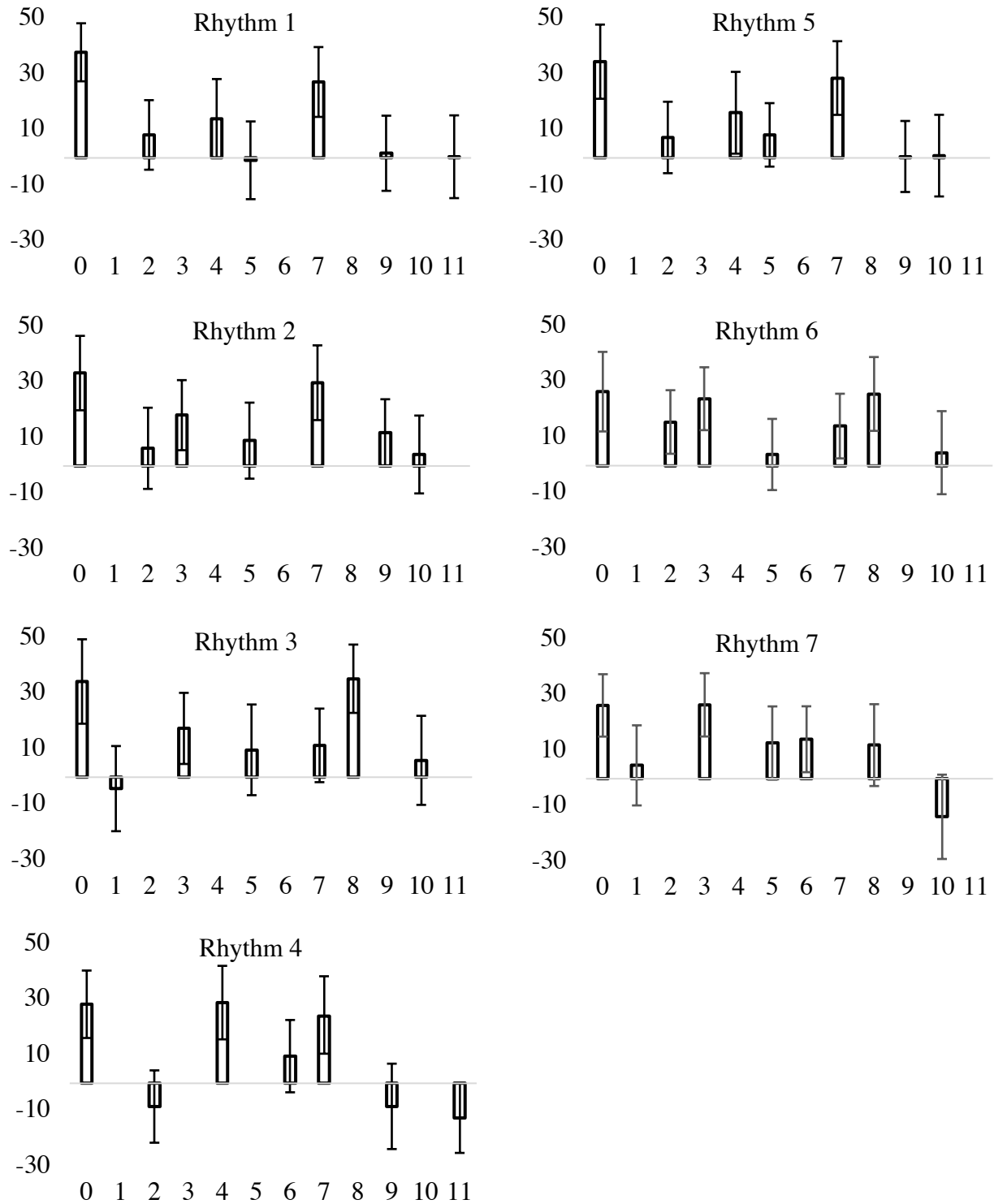


Figure 2.7. Probe accent ratings of each rhythm (y-axis; full range [-100, 100]) in the standard pattern labelled by clock position (x-axis). Error bars indicate 95% confidence interval.

that they perceptually divide into three groups with the accent preferred on the first event in each group. Thus, the preferred grouping appears to be 2 + 2 + 3 events, except when the fifth event is a short tone followed by a long tone, then the preferred grouping appears to be 2 + 3 + 2. In the statistical analysis, serial position and rhythm were coded as categorical variables, and participants as a random effect. A mixed-model multiple regression on the probe accent data with serial position, rhythm, and their interaction as factors was significant ($F(92, 2107) = 6.13$, $p < .0001$), and only serial position and the interaction between serial position and rhythm were significant ($F(6, 2107) = 23.68$, $p < .0001$ and $F(36, 2107) = 1.92$, $p = .0009$, respectively; rhythm: $F(6, 2107) = 1.31$, $p = .25$). A post-hoc Tukey HSD test showed the highest ratings overall were for accents in serial positions 1, 3, 5; the interaction reflects the shift to serial position 6 for R3 and R6.

Table 2.5 lists the remaining features that were tested, and the second column details their numerical coding. Tone duration codes whether the accented tone was short (eighth note) or long (quarter note). Metrical hierarchies were coded by assigning the largest number (4) to the strongest metrical level and the smallest number (1) to the weakest metrical level, with intermediate levels (3 and 2) in between. For example, Figure 2.5 shows the coding of the 3/2 metrical hierarchy (4 1 2 1 3 1 2 1 3 1 2 1) as the numbers in parentheses. These were then assigned according to when events occurred in the rhythms. For R1, for example, the seven events in the rhythm were thus coded 4 2 3 1 1 1 1. Implicitly, this assumes that the first event is the downbeat. An alternative would be to substitute 3 for 4 in each code, for example, 3 1 2 1 3 1 2 1 3 1 2 1 for 3/2. Codes for the syncopation-shifted metrical hierarchy were derived from the codes for metrical hierarchy as described above. The bolded temporal positions in Table 2.5

indicate the positions that would inherit the larger weight from the following temporal position when no event occurs there (as seen in Figure 2.5).

Table 2.5. *Coding and results from the mixed-model linear regression analysis on features of the standard pattern.*

Type of feature	Coding details	Statistical Test
Tone duration	1 = short, 2 = long	$F(1, 2154) = 1.92, p = .1656$
Metrical hierarchy	12/8 4 1 1 2 1 1 3 1 1 2 1 1	$F(1, 2154) = 53.31, p < .0001$
	6/4 4 1 2 1 2 1 3 1 2 1 2 1	$F(1, 2154) = 45.79, p < .0001$
	3/2 4 1 2 1 3 1 2 1 3 1 2 1	$F(1, 2154) = 57.88, p < .0001$
Syncopation- shifted metrical hierarchy	12/8 4 1 1 2 1 1 3 1 1 2 1 1	$F(1, 2154) = 29.94, p < .0001$
	6/4 4 1 2 1 2 1 3 1 2 1 2 1	$F(1, 2154) = 25.41, p < .0001$
	3/2 4 1 2 1 3 1 2 1 3 1 2 1	$F(1, 2154) = 135.85, p < .0001$

We conducted a separate mixed-model linear regression analysis of the probe accent ratings with each coded feature, with the results shown in the third column of Table 2.5. As can be seen, syncopation-shifted 3/2 metrical hierarchy was by far the strongest feature, and the only nonsignificant feature was tone duration. An additional analysis was conducted to test whether any of the other metrical coding contributed independently of the syncopation-shifted 3/2 metrical hierarchy. When it was entered into a step-wise regression with all the five other coded metrical hierarchies, it was the only feature that remained significant ($F(1, 2154) = 135.85, p$

< .0001). The alternative coding not assuming the first event was the downbeat resulted in a weaker effect. Two rhythms (R3 and R7) that start short-long might suggest an upbeat-downbeat pattern. However, note that for these rhythms, ratings of the accent on the first event are considerably higher than the second event.

To further examine the generality of the syncopation-shifted 3/2 metrical hierarchy across the rhythms, it was tested for each rhythm separately. The statistical result was significant for all seven rhythms individually (at $p < .005$). However, the syncopation-shifted 3/2 metrical hierarchy tended to fit the data best for those rhythms (R1, R4, R5) in which no syncopation shift had occurred in the first half of the rhythm. In these cases, events had occurred on strong metrical positions 0 and 4. The syncopation-shifted 3/2 metrical hierarchy fits the data less well for those rhythms (R2, R3, R6, R7) in which a syncopation shift occurred at clock position 3. This suggests that a longer and more regular rhythmic context preceding the syncopation shift makes it more prominent.

One question to consider is to what extent the probe ratings of individual participants varied under the syncopation-shifted 3/2 metrical hierarchy. To test this, we conducted a mixed-model multiple regression analysis on the probe accent data with participant (random factor), syncopation-shifted 3/2 metrical hierarchy (fixed factor), and the interaction between participant and syncopation-shifted 3/2 metrical hierarchy. The result was significant ($F(89, 2110) = 7.41, p < .0001$) for all three factors (syncopation-shifted 3/2 metrical hierarchy: $F(1, 2110) = 144.33, p < .0001$; participant: $F(44, 2110) = 8.06, p < .0001$; interaction: $F(44, 2110) = 3.65, p < .0001$).

A related question is whether or not familiarity, demographics, and musical training might explain the individual variation. To test this, we took the slopes of individual participants generated from the above-mentioned regression, and correlated it with every quantifiable

measure of familiarity, demographics, and musical training. These measures include average familiarity ratings of the seven standard rhythms, gender, age, whether or not the participant played an instrument, the number of years of instrumental or vocal training, whether or not they played piano, whether or not they took a class that included music theory, whether or not they took a class that mentioned African rhythm, whether or not they listen to classical music, and their familiarity with Afro-Caribbean music, Jazz, Latin dance, or African dance. No significant correlation was found (p ranged between .11 and .83). This suggests that familiarity, demographics, and musical training do not explain the individual variations in the probe accent ratings predicted by the syncopation-shifted 3/2 metrical hierarchy.

Distances between rhythms

To quantify the perceived distances among the seven rhythms, we calculated city-block distance. The city-block distance between any two rhythms was calculated as the sum of the absolute difference in probe ratings at each corresponding serial position. For example, to calculate city-block distance between R1 and R2, we found the absolute difference of the ratings of accents on the first event in R1 and R2, repeated this for the accents on the other six events, and then summed the seven absolute values. This creates a matrix of distances between all possible pairs of rhythms. It is worth noting that events at the same serial position in different rhythms do not necessarily take place at the same clock position. In other words, clock position is not taken into account in the calculation of city-block distance, only serial position was considered.

This matrix was then correlated with the two theoretical measures of distances between rhythms, which are exchanging “halves” (see Table 2.2) and swap distance (see Table 2.4). Exchanging “halves” ($R^2 = .24, F(1, 19) = 5.85, p < .05$) and swap distance ($R^2 = .24, F(1, 19) =$

6.07, $p < .05$) accounted for an equal amount of the variation in city-block distances. Both measures order the rhythms as R4, R1, R5, R2, R6, R3, and R7, but the first is a circular representation and the second is a linear representation. To examine this in more detail, the city-block distances were plotted against swap distance and the result was somewhat curvilinear. Supporting this, a polynomial regression to degree 2 provided a better account of the city-block distances ($R^2 = .33$, $F(2, 18) = 4.34$, $p < .05$) than swap distance. Thus, R4 and R7 were closer than predicted by swap distance but not as close as in the circular representation of exchanging “halves.”

Discussion

The probe accent ratings for the seven rhythms showed two perceptual influences. The first related to how the events were organized into subgroups. The rhythms tended to subdivide into three subgroups that were of different durations. These subgroups could be understood in terms of a number of underlying perceptual principles that contributed to grouping. First, the relatively high ratings for accents in serial position 1 suggest that listeners start with a preference for an accent on the first event. This was true regardless of whether the first event was short or long. Secondly, the relatively high ratings on serial position 3 suggests listeners prefer groups of two. Previous research has found a general preference for duple grouping in isochronous sequences (Brochard, Abecasis, Potter, Ragot, & Drake, 2003; Martens & Benadon, 2017). These data suggest that the preference for duple grouping might extend also to non-isochronous rhythms. Alternatively, or in addition, this result might reflect a preference for alternating between contrasting levels of stress, perhaps analogously to speech. These two preferences provide evidence for Locke’s (1982) idea that both identifying the cycle and hearing the accents within that cycle are important cognitive principles for the standard rhythms.

There were two rhythms (R3 and R6) where the second sub-group contained three rather than two events and the preferred accent occurs on serial position 6 instead. These are both cases where a short event in serial position 5 precedes a long event in serial position 6. Thus, a preference for an accent on a long event following a short event can override the preference for groups of two. In addition, accents were never preferred in serial position 7, the penultimate tone, with the stress likely accruing to the immediately following final tone. The subgroups of 2 + 2 + 3 and 2 + 3 + 2 together suggest a preference for balanced grouping, where each subgroup consists of a roughly equal number of events (Handel & Todd, 1981). Finally, the highest familiarity and well-formedness ratings of R4 suggest the preference for beginning a rhythmic sequence with the longest run of identical events (the same tone duration). This run principle is consistent with early evidence on the perceptual organization of temporal patterns (Handel, 1974; Preusser, Garner, & Gottwald, 1970).

The second perceptual influence apparent in the results is described by the syncopation-shifted 3/2 metrical hierarchy proposed by Temperley (1999, 2000). It modifies the metrical hierarchy for 3/2 to accommodate the effect of syncopation, which also provides evidence for Locke's (1982) observation of shifting accents. When an event is not sounded at a time point that would normally be heard as metrically strong, its metrical weight transfers to the immediately preceding time point where an event is sounded. This theoretical proposal provided a much better account than any of the other metrical hierarchies that were tested. It also fits the data well across each of the seven rhythms individually. However, there was a slight tendency to fit the data better for rhythms in which the first syncopation shift occurred later in the sequence. These began with a longer, more regular beginning which may have contributed to the salience of the later syncopation shift. To sum up, the probe accent results showed hierarchies that depended on

two types of influences: those following principles of grouping, and those following metrical principles modified to account for the effect of syncopation. Thus, as predicted by Locke (1982) and Pressing (1983), these perceptual principles appear to work together despite the complex relationship between grouping and meter.

Finally, the probe accent data also tested two theoretical measures of distances between rhythms: exchanging “halves” and swap distance. Both order the rhythms the same way, except that exchanging “halves” is a circular model and swap distance is a linear model. The distance measure derived from the empirical data was a measure of the similarity of the probe accent ratings for all pairs of rhythms. Two rhythms were considered close if the ratings of the accents in the corresponding seven serial positions were similar. These distances derived from the data ordered the rhythms in the same way as both exchanging “halves” and swap distance, with a curvilinear component that suggested a model lying somewhat between the circular and linear models.

CHAPTER 3

EXPERIMENT 2

Pressing (1983) noted that the diatonic scale pattern, which is the most prominent scale in Western tonal music, is isomorphic to the standard rhythmic pattern in Experiment 1. Scales with the diatonic pattern can be characterized by two common structural properties (Agmon, 1989): their *scale steps* (i.e., 12 for the diatonic pattern) and the *diatonic intervals* in the cycle (i.e., 7). The scale in the form of C D E F G A B is referred to as the parent scale (Rechberger, 2008), and the other six related scales in the diatonic family are traditionally viewed as modes. The diatonic pattern is *rotational*, such that the seven starting points generate seven unique patterns (see the last three columns in Table 2.1). Figure 2.4 illustrates the same scales in terms of tone identity (C, Db, D, etc.).

According to Powers (2001), mode applies to three separate stages of Western music: Gregorian chant, Renaissance polyphony, and tonal harmonic music of the 17th century to the 19th century. As a theoretical concept, it was originally concerned with classifying scales according to such elements as the final scale degree, and the first and the fourth scale degrees. In the 20th century, in connection with music of other cultures, the term mode has broadened to include melodic and motivic features. In the present context, we use mode in the first sense, that is, a scale beginning and ending on a certain tone and containing a specific pattern of intervals between tones.

In this section, we will briefly describe three types of features that will be considered in analyzing the probe tone data. We call the first type *basic* features, and they are shown in Table 3.1. The first basic feature is the interval size between each tone and the next tone as measured in semitones. In the experiment, we used both ascending and descending scales. For example, for

Table 3.1. *Coding and results from the mixed-model linear regression analysis on basic features of the diatonic pattern.*

Basic features	Coding details	Statistical Test
Interval size	1 = semitone, 2 = whole tone to the following pitch	$F(1, 4363) = 11.33, p = .0008$
Primacy	1, 2, 3, 4, 5, 6, 7 when ascending, 1, 7, 6, 5, 4, 3, 2 when descending, corresponding to C, Db/D, Eb/E, F/F#, Gb/G, Ab/A, Bb/B	$F(1, 4363) = 521.51, p < .0001$
Chroma distance from C	0, 1, 2, 3, 4, 5, 6 corresponding to C, Db/B, D/Bb, Eb/A, E/Ab, F/G, F#/Gb	$F(1, 4363) = 230.12, p < .0001$
Steps on circle of fifths from C	0, 1, 2, 3, 4, 5, 6 corresponding to C, F/G, D/Bb, Eb/A, E/Ab, Db/B, F#/Gb	$F(1, 4363) = 395.71, p < .0001$

the ascending scale C D E F G A B C, C was coded as 2, D was coded as 2, E was coded as 1, and so on; for the descending scale C B A G F E D C, C was coded as 1, D was coded as 2, E was coded as 2, and so on. Three other features are coded with respect to the tone C. The first is primacy, which codes how early in the context sequence the probe tone occurred after C. The second is distance from C on the chroma circle; the third is distance from C on the circle of fifths. Krumhansl (1990) reported an experiment in which a major or minor context was followed by two tones and listeners rated how well the second followed the first. The judgments were influenced by the distance between the tones on the chroma circle and the circle of fifths, which suggested including these variables in the analyses.

The second type is called *tonal* features, the first of which is consonance between the probe tone and the tone C. Krumhansl (1990, Table 3.1) lists six different consonance measures that

were obtained perceptually or computationally. Three more recent measures were also considered. McDermott, Lehr, and Oxenham (2010) obtained pleasantness ratings of two-tone chords in four different timbres, out of which, the synthetic tone in their Figure 2.1 is most similar to the stimuli used in the current study. Stolzenburg (2015, see Table 3) computed smoothed relative periodicity of dyads. Bowling, Purves, and Gill (2018) obtained dyad ratings of consonance, defined as pleasantness or attractiveness, rank ordered in their Figure 2.1. Other tonal features included Large's (2010) oscillatory neurodynamic model with the values, computed as in Large, Kim, Flaig, Bharucha, and Krumhansl (2016), and Gill and Purves's (2009) harmonicity values. The final three measures come from the study by Krumhansl and Kessler (1982): C major, C minor, and the major profile for each mode corresponding to the number of accidentals (C major for M1, Bb major for M2, Ab major for M3, G major for M4, F major for M5, Eb major for M6, and Db major for M7).

The third type is *modal* features. Figure 3 of Huron and Veltman (2006) shows the frequency of occurrence for each pitch class in church modes in a random sample of 98 Gregorian chants. The samples were larger for some modes than others, so we converted the values to percentages. Out of the 8 mode profiles, the Dorian and the Hypomixolydian modes can both correspond to M2, so we considered both scenarios.

As with the rhythms, there are two representations of the distances between modes. A circular representation can be derived by analogy with *exchanging "halves"* (Agawu, 2006; see Table 2.2 and Figure 3.1). Each mode can be divided into two "halves" where one consists of two whole tones and one semitone (5 semitones), and the other of three whole tones and one semitone (7 semitones). The bottom row and the last column of Table 2.2 show the distances between the seven modes in terms of exchanging "halves." Second, a linear representation of

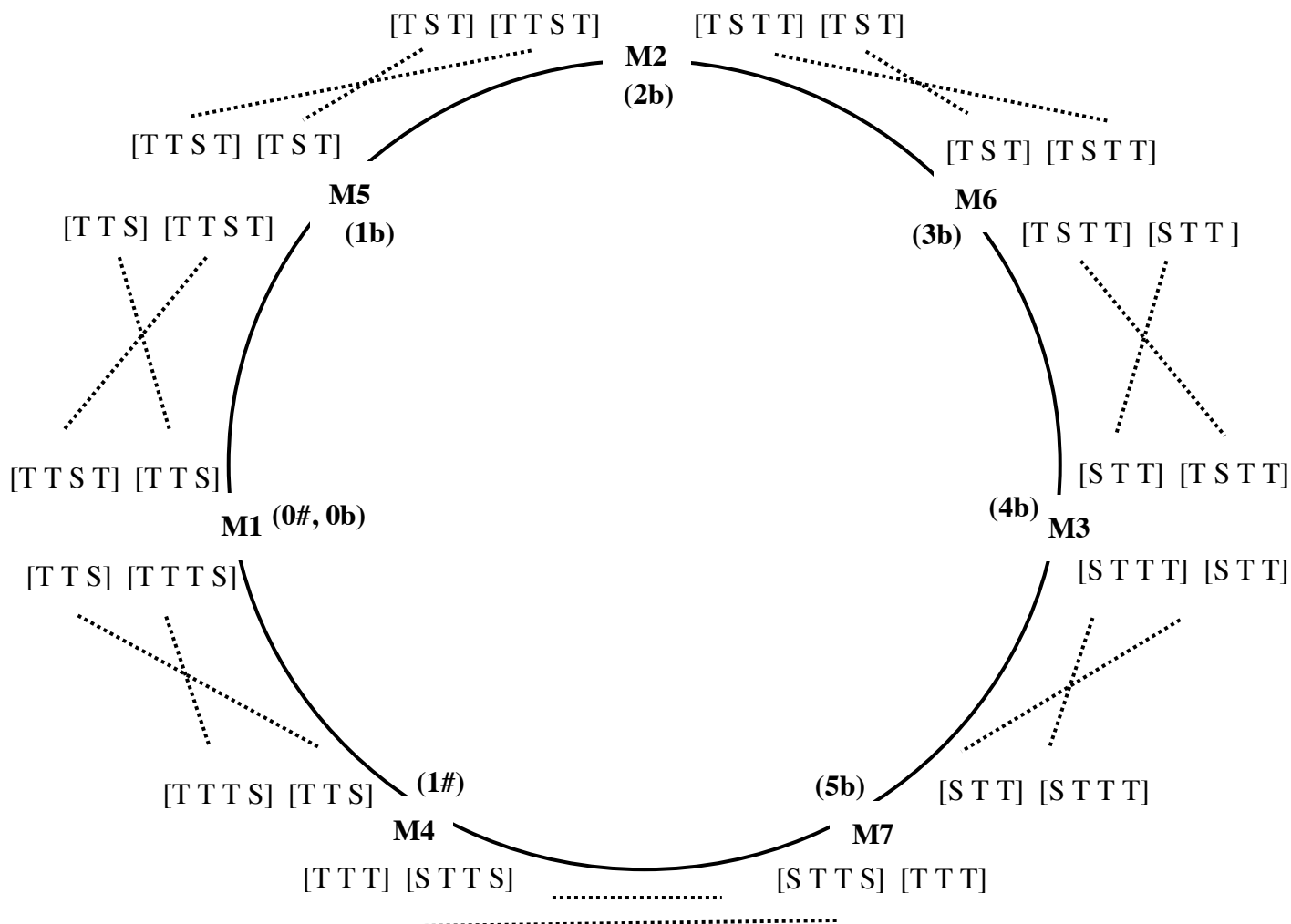


Figure 3.1. Exchanging “halves” for the diatonic modes. The seven modes are arranged in the circular order of M4, M1, M5, M2, M6, M3, M7. The square brackets arrange the structure of each rotation into two types of approximate “halves” (5 + 7 semitones and 7 + 5 semitones). The dotted lines mark the exchanges between “halves” to transform one rotation into a neighboring rotation.

mode distances results from the swap process illustrated in Table 2.3. The resulting distance values are shown in Table 2.4. The ordering in both representations corresponds to the number of

sharps and flats. However, because M6 can also be considered as the parallel minor of M1, M6 might better be placed between M1 and M5.

It might be of interest that swap distance is related to a way of differentiating between modes (Curtis, 1992; Power, 2001). Table 3.2a shows the three types of species of the 4th (3 intervals covering a span of a fourth). Table 3.2b shows the four types of species of the 5th (4 intervals covering a span of a fifth). Table 3.2c shows how each mode can be expressed as one species of the 4th and one species of the 5th. The ordering in swap distance can be derived by identifying the overlapping species, highlighted in gray.

Finally, modes are often differentiated based on modal ethos, or emotion. Powers (2001, Table 10) summarized Zarlino's modal ethos of the twelve Glarean's modes, which correspond as follows: M1 = 11, M2 = 1, M3 = 3, M4 = 5, M5 = 7, M6 = 9, M7 = 4. In the order M4, M1, M5, M2, M6, M3, and M7, modes toward one extreme have positive ethos (M4 = "joyous, modest and pleasing;" M1 = "suitable for dances, lascivious") and modes toward the other extreme have negative ethos (M3 = "somewhat hard, moves one to weeping;" M7 = "lamenting sad, supplicant lamentation"). An interesting exception is that M6 is relatively positive ("cheerful, sweet, soft and sonorous"), placing it nearer to M1, its parallel major.

Experimentally, Temperley and Tan (2013) showed that the emotion connotation of modes in the diatonic pattern ranged from the happiest to the sadist in the same order (M7 was not tested), except M1 was judged happiest, likely due to familiarity. Both the theoretical modal ethos and the empirical emotion connotations suggest that the "sharpness" of a mode correlated with perceived happiness.

Objectives of the Study

Experiment 2 was designed to determine the tonal hierarchies of the modes using the probe

Table 3.2. *Species of the 4th and species of the 5th in the diatonic pattern.*

a) Types of Species of the 4th

Type	Order of pitch intervals
1	T-S-T
2	S-T-T
3	T-T-S

Note. T = whole tone, S = semitone.

b) Types of Species of the 5th

Type	Order of pitch intervals
1	T-S-T-T
2	S-T-T-T
3	T-T-T-S
4	T-T-S-T

Note. T = whole tone, S = semitone.

c) Possible combinations for each mode

Species of the 4th + Species of the 5th			Species of the 5th + Species of the 4th		
M4(1#)	n.a.		T-T-T-S (type 3)	+	T-T-S (type 3)
M1 (0#, 0b)	T-T-S (type 3)	+	T-T-T-S (type 3)		
			T-T-S-T (type 4)	+	T-T-S (type 3)
M5(1b)	T-T-S (type 3)	+	T-T-S-T (type 4)		
			T-T-S-T (type 4)	+	T-S-T (type 1)
M2(2b)	T-S-T (type 1)	+	T-T-S-T (type 4)		
			T-S-T-T (type 1)	+	T-S-T (type 1)
M6(3b)	T-S-T (type 1)	+	T-S-T-T (type 1)		
			T-S-T-T (type 1)	+	S-T-T (type 2)
M3(4b)	S-T-T (type 2)	+	T-S-T-T (type 1)		
			S-T-T-T (type 2)	+	S-T-T (type 2)
M7(5b)	S-T-T (type 2)	+	S-T-T-T (type 2)		
				n.a.	

Note. T = whole tone, S = semitone, n.a. = not applicable.

tone technique (Krumhansl & Shepard, 1979), in which a probe tone is played after the mode context. Listeners rated how well the probe tone fits into the context using the seven pitches in each mode. Modes were presented in both ascending and descending forms. Using the probe tone

method, the current study will obtain a profile of perceptual stability for all pitches in each of the seven modes. These data will be analyzed for the effects of the features described above: basic, tonal, and modal. The data will also be used to test the distance representations.

Method

Participants

Forty-five students (6 males and 39 females, age 18-40, 31 also participated in Experiment 1) participated in Experiment 2 for course credit or \$5 cash. On average, they have played music for 14.0 years (including music lessons; summed over all instruments and voice; range, 1-29; three had absolute pitch, and one had no musical training). The study was approved by Cornell University's Institutional Review Board.

Apparatus and Stimulus Materials

We constructed seven modes based on the major diatonic mode by shifting the starting pitch interval (see Figure 2.4 and Table 2.1). All modes consisted of 8 isochronous tones, each lasting for a quarter note, forming 7 pitch intervals. Each mode was constructed in both ascending and descending forms, beginning and ending on C. The range of the ascending sequence was C3 to C4, and the range of the descending sequence was C6 to C5 (as in Krumhansl & Shepard, 1979). The seven modes were followed by a probe tone, played in the range of Db4 to C5 for the ascending mode and C4 to B4 for the descending mode. Only the seven tones in each mode were used as probe tones. The pause of silence in between the mode and the probe tone lasted for two quarter notes. All mode contexts were at the tempo of 140 quarter-note per minute (each quarter note lasts approximately 440 ms; each trial lasted approximately 4.84 s) and was sounded in piano timbre. Also constructed were 7 modes without a probe tone after the context, called the *neutral* modes.

Procedure

Participants rated how well the probe tone fits into the scale by moving a slider on a continuous scale ranging from extremely bad fit (left end) to extremely good fit (right end). They began with four practice trials. Next, they listened to each of the 98 probe tone trials (49 ascending and 49 descending), which were blocked by mode type (i.e., M1, M2, etc.); the ascending and descending modes were intermixed. The 7 blocks were presented in a randomized order. Each block began with the neutral mode played twice, once ascending and once descending, to familiarize the listener with the mode to be presented in the block. This was followed by 14 probe tone trials of the same mode played in randomized order. Following the probe tone ratings, participants listened to each of the 7 neutral modes (without the probe tone) in ascending form in a randomized order and rated them on two scales: 1) how familiar the scale was prior to the experiment, and 2) how well-formed the scale seems, in other words, whether or not the scale forms a good pattern. At the end of the study, they filled out a demographics questionnaire. The experiment lasted approximately 30 minutes.

Results

All continuous rating scales were coded from -100 to 100. Ratings from 91% of participants correlated significantly with the average probe ratings, suggesting considerable agreement across participants. However, a clustering analysis (Ward method) was done to see if there were identifiable demographic differences between the three clusters that the clustering analysis found. Demographic factors included whether they had absolute pitch, whether they have ever studied instruments or voice, years of musical training, whether on piano or not, music theory class, whether church modes were mentioned in a class they had taken, and whether or not they listen to classical music. Of these, only one factor differed between the groups and the difference

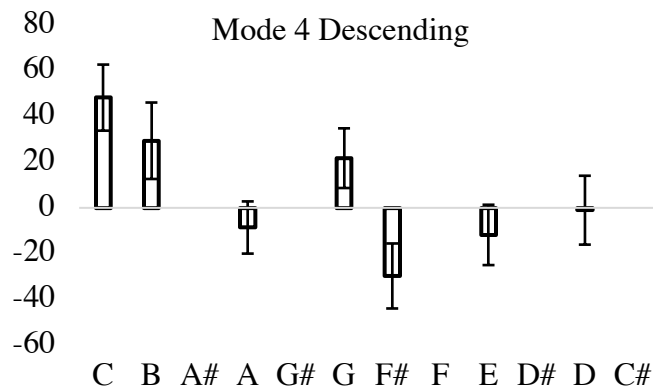
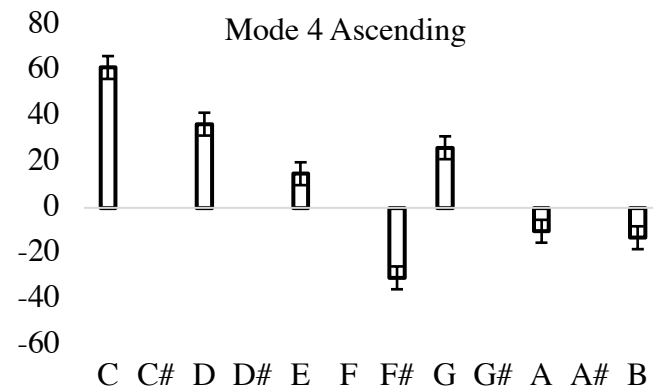
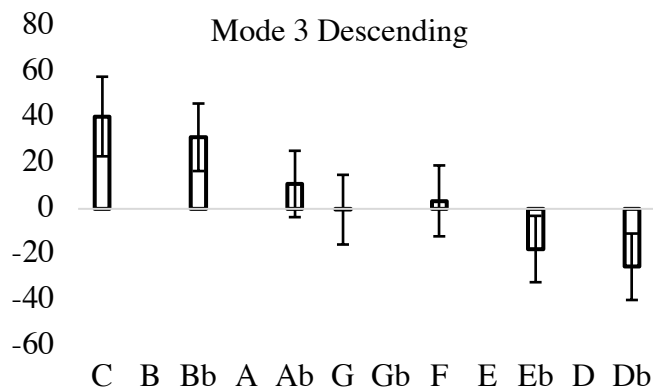
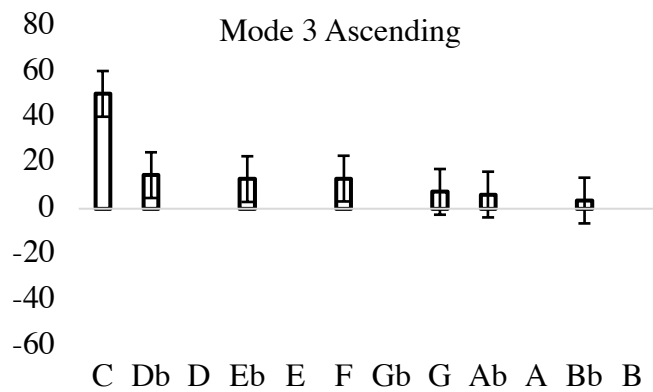
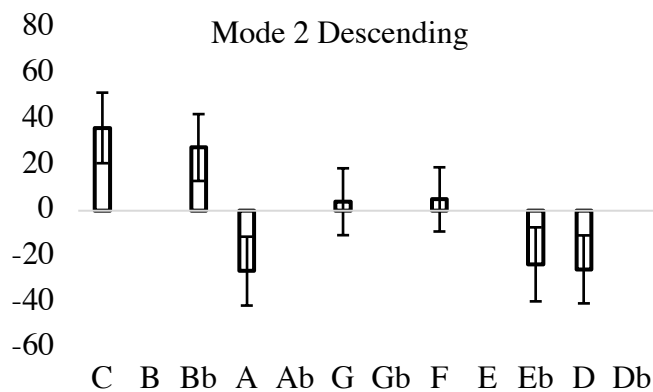
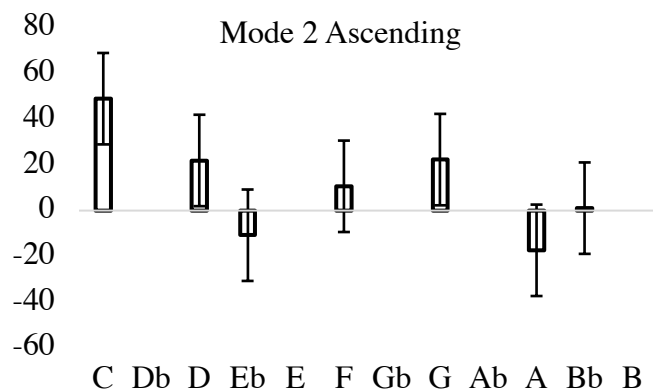
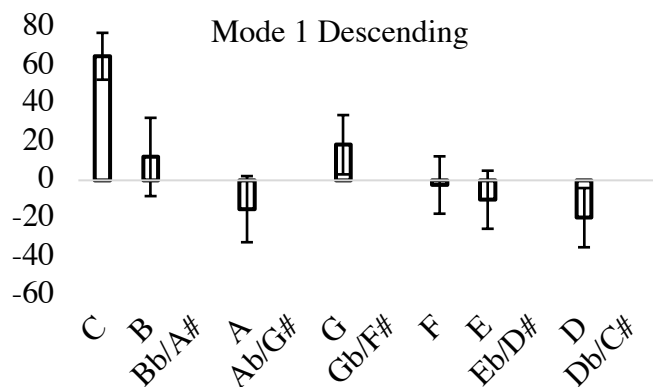
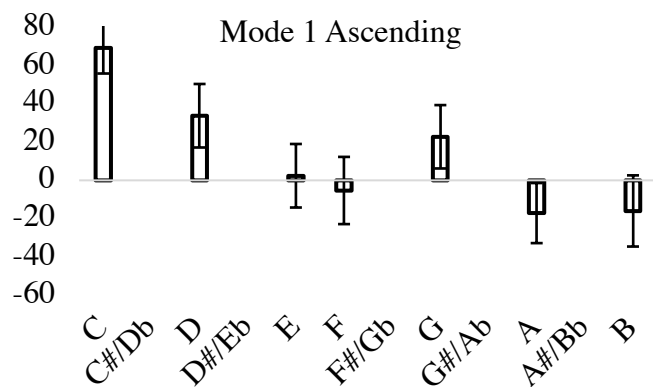
between the groups was only weakly significant (whether or not they had studied a musical instrument or voice: $F(2, 41) = 3.34, p = .05$).

A related question is whether some modes were more familiar than others prior to the experiment. To test this, we conducted a mixed-model regression analysis and found some modes were more familiar than others ($F(6, 263) = 16.67, p < .0001$). A post-hoc test (Tukey HSD) showed that M1, the major diatonic scale, was rated as the most familiar, M7 was rated as the least familiar; the others were not different statistically. The well-formedness ratings were similar, except that M2 was not statistically different from M7. The familiarity and well-formedness ratings correlated significantly with each other ($F(1, 268) = 149.93, p < .0001$).

Hierarchy within a mode

Figure 3.2 shows the probe tone ratings of each of the seven modes plotted according to the tone identity of the probe tone and whether the context was ascending or descending. In the mixed-model multiple regression with mode and tone identity as fixed effects and participants as a random effect, tone identity was highly significant ($F(12, 4352) = 27.16, p < .0001$). A post-hoc Tukey HSD test showed C was higher than all other tones; it was followed by G. Then the test identified a group containing Db, D, F, Ab, Bb, and B which were not significantly different from each other. With the exception of Ab, these tones are either proximate to C (Db, D, Bb, and B) or close to it on the circle of fifths and also consonant with it (F). The tritone, F#/Gb, always received the lowest ratings when it was present (M4 and M7).

Basic features. We first conducted a separate mixed-model linear regression analysis on each basic feature shown in Table 3.1. As can be seen, all these variables were statistically significant. Primacy was a very strong predictor of the probe tone ratings. That is, listeners preferred probe tones that matched the chroma of tones sounded early in the context. In general,



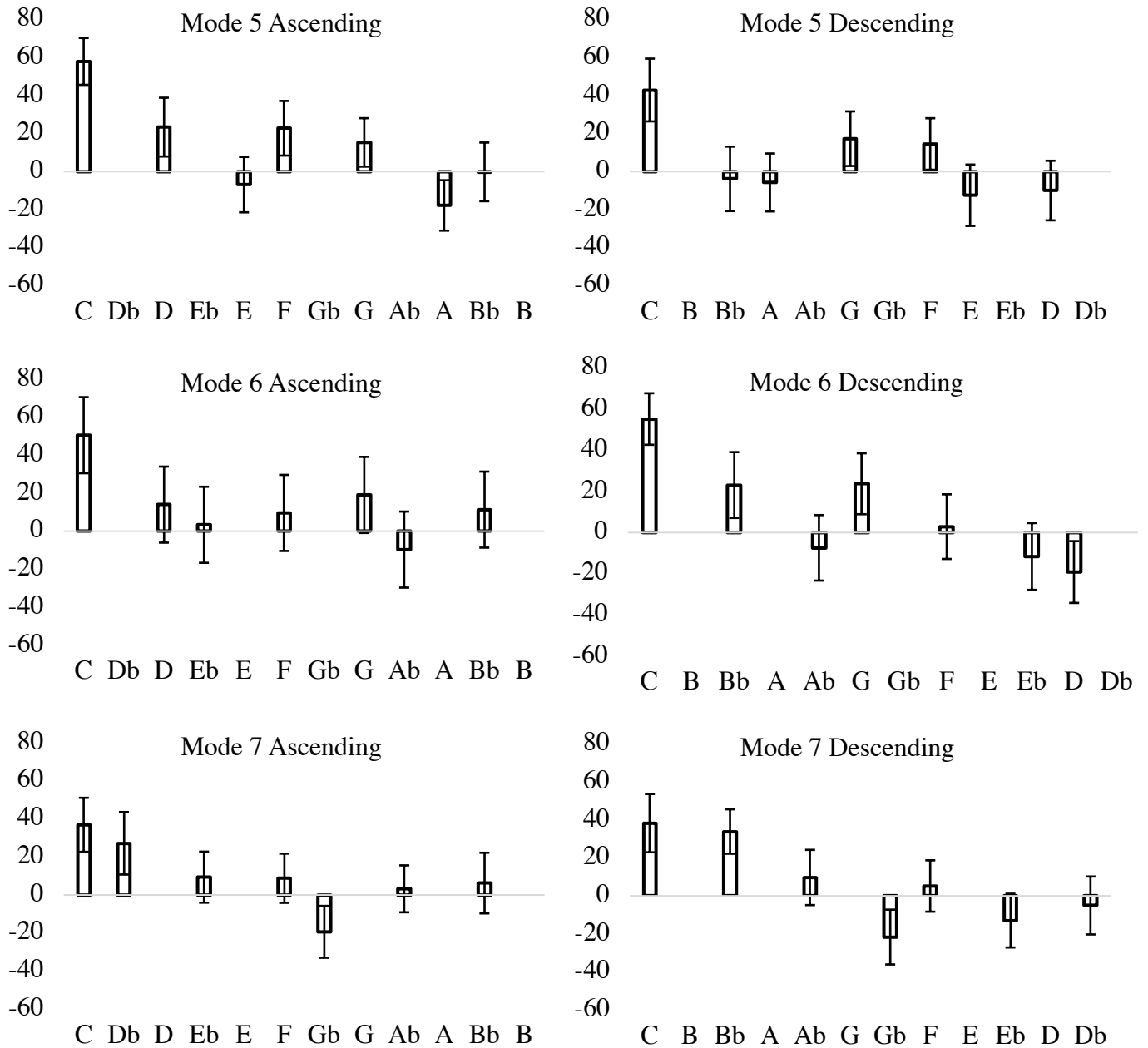


Figure 3.2. Probe tone ratings of each mode (y-axis; full range [-100, 100]) in the diatonic pattern labelled by tone identity (x-axis), with ascending on the left and descending on the right. Error bars indicate 95% confidence interval.

listeners preferred tones that are close to the C on the circle of fifths, and to a lesser extent, on the chroma circle, over those that are farther from the C. These results support the idea that the

C, which is sounded first and last in the context, serves as a reference point. Interval size had a relatively weak effect. When all the basic features were entered into one regression model, interval size was only marginally significant.

Consonance, tonal, and modal features. All the consonance measures correlated significantly with the probe tone ratings (at $p < .0001$) except for the values of McDermott et al. (2010) which had no value for the tone C. This suggested that the significant correlations for the other consonance measures might be largely driven by the high ratings of the tone C. When the C was removed from the statistical analysis, only two remained significant: the values from Kameoka and Kuriyagawa (1969, Fig. 8) and Helmholtz Equal-Temperament (1885/1954, p. 332). Other features were also correlated (at $p < .0001$) with the probe tone data: the prediction of the neurodynamic model (Large, 2010), harmonicity (Gill & Purves, 2009), the C major and C minor probe tone profiles (Krumhansl & Kessler, 1982, see Figure 1.4), and the profiles of the keys corresponding to the number of accidentals (whether M6 was major or minor). When C was excluded from the analysis, the predictions of the neurodynamic model were no longer significant, but the other predictors remained significant (at $p < .01$). Finally, the distribution of tones computed as percentages in Huron and Veltman's (2006) Gregorian chant corpus did not significantly correlate with the probe tone data assuming either M2 = Dorian ($F(1, 4363) = .0447, p = .8325$) or M2 = Hypomixolydian ($F(1, 4363) = .2021, p = .6531$).

Distances between modes. The first approach taken to quantifying the perceived distances between the seven modes was to calculate city-block distance in the same way as for the seven rhythms in Experiment 1. C received the highest rating for all the modes, so we considered it to be the first scale degree for each mode, with the remaining tones considered scale degrees 2, 3, and so on. In the context of modes, city-block distance is the sum of the absolute difference in

probe tone ratings for each scale degree. This creates a matrix of distances between all possible pairs of modes, which correlated significantly with swap distance ($R^2 = .72$, $F(1, 19) = 47.69$, $p < .0001$), but only marginally with exchanging “halves” ($R^2 = .15$, $F(1, 19) = 3.43$, $p = .0798$). This result suggests that the distances between modes, as measured by the similarity of their probe tone ratings, can be quite well accounted for by swap distance.

The approach just described collapsed different tones into the same category, for example, D and Db, so a second approach was taken. It uses the probe tone profiles from Krumhansl and Kessler (1982) for C major, C minor, and the keys corresponding to the number of accidentals with M6 coded as either major or minor. To assess the relative strengths of these four tonality predictors, they were all entered into a step-wise regression, together with the basic factors already found to be significant (proximity to the beginning, chroma distance from C, distance on the circle of fifths from C). The three basic features remained significant. The most significant of the probe tone profiles was the one with the corresponding majors (except M6 coded as C minor). With this exception, this is the same ordering as swap distance (and corresponds to the number of sharps and flats). So, this approach again supports swap distance as the best measure of inter-mode distance.

Discussion

The probe tone ratings for the seven modes showed a number of underlying perceptual influences. The first was how early the probe tone was sounded in the context, the second was proximity to C on the chroma circle, and the third was proximity to C on the circle of fifths. On balance, consonance had a relatively weak effect on the probe tone ratings, but stronger effects were found for harmonicity (Gill & Purves, 2009) and tonal hierarchies (Krumhansl & Kessler, 1982). The mixed major/minor probe tone profiles (with M6 considered C minor and the other

the major key with the corresponding key signatures) was the strongest of the tonal profiles. These results suggest listeners were basing their judgments in part by reference to tonal hierarchies in the more familiar major and minor keys. Lastly, we also found no significant relationship between the probe tone ratings and the distribution of tones in Huron and Veltman's (2006) corpus study of Gregorian chants.

The measure of inter-mode distance derived from the probe tone hierarchies supported the swap distance measure, which also orders the modes according to the number of accidentals (and as on the circle of fifths). The circular exchanging "halves" distance was less well supported. The order was also supported by the ordering as on the circle of fifths based on the probe tone profiles for keys, although as just noted M6 was closer to C major, its parallel major. Finally, the order generally fits with the ordering of modes in terms of positive to negative affect (Powers, 2001; Temperley & Tan, 2013).

CHAPTER 4

EXPERIMENT 3

A basic question in music cognition is how the dimensions of pitch and time combine. The literature has found two different kinds of relationships (as summarized in Krumhansl, 2000; Prince 2011; Prince & Schmuckler, 2014). One is that pitch and time are two separable dimensions, where changes in one dimension do not affect the perception of the other dimension. The other is that pitch and time interact perceptually, where changes in one dimension affect the perception of the other dimension. The current experiment is concerned with the effect of the alignment between pitch patterns and rhythmic patterns on the perception of the combined pattern, specifically, whether the rhythm is heard as fitting well with the mode. We will briefly review studies that have focused on combined pitch-time patterns in particular and explored the various factors that seem to promote one type of processing over the other.

A number of studies have supported the idea that pitch and time are two separable dimensions. An early study found pitch and time made independent contributions to melodic similarity judgments, with trade-offs between rhythmic and pitch factors (Monahan & Carterette, 1985). Palmer and Krumhansl (1987a, 1987b) investigated how pitch and time influence judgments of phrase endings by shifting the melody relative to the rhythm. The judgments were consistent with an additive model of the final tone's position in tonal and metrical hierarchies. Prince and Pfordresher (2012) used combinations of rhythmic and pitch patterns that varied in complexity by manipulating the number of distinct durations and pitches. The numbers independently influenced complexity judgments. Prince (2014) investigated how surface and structural properties of pitch and time affect the perceived similarity of melodies. When manipulated independently, they made independent contributions to similarity judgments.

Other studies, however, demonstrate interactions between pitch and time. Deutsch (1980) found more accurate dictation when the groups in the pitch pattern coincided with the groups in the temporal pattern. Jones, Summerell, and Marshburn (1987) demonstrated better recognition of melodies when played in the original rhythms than in a novel rhythm. Boltz (1995) presented a set of folk tunes in which tones were prolonged either at phrase endings or somewhere else within the phrase. For the latter, duration judgments were less accurate. Boltz (1998) found similar effects for tempo judgments. Prince (2011) asked listeners to selectively attend to pitch (i.e., tonality), time (i.e., meter), or both, and judge melodic goodness. They were unable to ignore the unattended dimension; interactions were strong when attending to both dimensions. Prince and Loo (2017) found similar failures in selective attention in expectancy ratings. Prince and Pfordresher's (2012) participants made more reproduction errors in one dimension when the other dimension was more complex.

Finally, a corpus study of compositions by Bach, Mozart, Beethoven, and Chopin showed that tonally stable pitch classes disproportionately fell on temporally stable positions (Prince & Schmuckler, 2014), that is, metrical and tonal hierarchies coincide. Similarly, corpus studies of bebop-styled jazz showed that improvisers emphasized the tonally prominent chords at metrically prominent temporal positions (Järvinen, 1995; Järvinen & Toiviainen, 2000). These findings suggest that the perceptual question concerning how pitch and time interact has significance for how these dimensions are combined in music.

Prince, Schmuckler, and Thompson (2009) and Prince (2011) summarized various factors that might explain these different results. One explanation concerns early versus late processing, where pitch and time are initially processed separately (e.g., pitch height and tempo) and are later integrated (e.g., melodic completion and similarity) (Hébert & Peretz, 1997; Peretz & Kolinsky,

1993). A second explanation is related to local versus global processing of the music. Experimental tasks that focus on a local or isolated event (e.g., a specific chord or rhythmic event) are likely processed independently whereas tasks that attend to a global event that spans across a larger time period (e.g., tension and relaxation within a sequence) require processing of integrated pitch and time information (e.g., Tillmann & Lebrun-Guillaud, 2006). A third explanation is the degree of perceived musical coherence of the musical context (Boltz, 1999). Pitch and temporal information tend to be processed in a unified way when a melodic sequence has high musical compatibility, where the pitch structure (e.g., phrase) and the temporal structure (e.g., rhythm) are consistent with each other (e.g., coincide). In contrast, dimensions in less coherent contexts (e.g., misaligned structures) might be processed separately.

Prince (2011) pointed to two other factors that might affect how pitch and time are processed. First, he distinguishes the differences between tasks that monitor relatively surface aspects of the music (e.g., contour or pattern of durations) and more abstract structures (e.g., tonal and metrical hierarchies). In the former case, the dimensions might be processed more independently than in the latter case. The second factor is the relative salience of the dimensions. If there is a tendency for one perceptual dimension to be more salient (i.e., pitch may be more prominent for Western listeners), it may dominate the other (i.e., time), resulting in an asymmetric interaction between pitch and time.

Objectives of the Study

Experiment 3 tests whether listeners prefer specific pitches to occur in specific temporal positions using the theoretically isomorphic standard pattern and diatonic patterns. Each rhythm was combined with each mode to create a total of 49 (7 x 7) combinations, and listeners judged how well the rhythm fits the scale. For brevity, we will call these *fit judgments*. Two possible

outcomes will be tested. The first is that listeners will prefer large pitch intervals to coincide with long tone durations, a surface-level match. The second possible outcome is that they will prefer tonally stable tones, as measured in Experiment 2, to coincide with temporal positions where accents were preferred, as measured in Experiment 1, a hierarchical-level match.

Method

Subjects

Fifty students (15 males and 35 females, age 18-40) participated in Experiment 3 for course credit. On average, they have played music for 11.3 years (including music lessons; summed over all instruments and voice; range, 1-40; three had absolute pitch, and four had no musical training). The study was approved by Cornell University's Institutional Review Board.

Apparatus and Stimulus Materials

We constructed forty-nine rhythm-mode pairs by combining each of the seven rhythms in Experiment 1 with each of the seven ascending modes in Experiment 2 (see Table 2.1). Descending modes were not used because the interval sizes would not then be isomorphic to the tone durations in the rhythms. The range of all sequences was C4 to C5. All stimuli were played in piano timbre at the tempo of 120 quarter-note per minute (each quarter note lasts approximately 500 ms; each trial lasts approximately 3.50 s). In addition, we used the seven unaccented rhythms and the seven modes without the probe tone as the *neutral* trials.

Procedure

Participants rated how well the rhythm fits the scale by moving a slider on a continuous scale from extremely bad fit (left end) to extremely good fit (right end). They began with four practice trials. Next, they listened to and rated each of the 49 trials in randomized order. After completing a demographics questionnaire, they then repeated the same task one more time, with a different

randomized order of trials. Following the fit judgment, participants completed another demographics questionnaire, and then listened to each of the 14 neutral trials (7 unaccented rhythms and 7 modes without the probe tone) in a randomized order and rated them on: 1) how familiar the rhythm or mode was prior to the experiment, and 2) how well-formed the rhythm or mode seems, in other words, whether or not it forms a good pattern. The experiment lasted approximately 30 minutes.

Results

All continuous rating scales were coded from -100 to 100. Ratings from 86% of participants correlated significantly with the average probe ratings, suggesting considerable agreement across participants. However, a clustering analysis (Ward method) was done to see if there were identifiable demographic differences between the two clusters that were found. Demographic factors included whether they had absolute pitch, whether they have ever studied instruments or voice, years of musical training, whether on piano or not, whether or not they listen to classical music, and their familiarity with music or dance related to the standard pattern (Afro-Caribbean music, Jazz, Latin Dance, and African Dance). Only three of these differed between the groups and the difference was only weakly significant (whether or not they have absolute pitch: $t(33) = -1.79, p = .04$; whether or not they had studied a musical instrument or voice: $t(33) = 2.10, p = .04$; whether or not they listen to classical music: $t(23) = 2.23, p = .04$).

As in Experiments 1 and 2, one question to consider is whether some rhythms or modes were more familiar than others prior to the experiment. A mixed-model regression analysis showed some rhythms were more familiar prior to the experiment than others ($F(6, 293) = 4.25, p = .0004$); a post-hoc test (Tukey HSD) found R7 was rated as the least familiar, but the others were not statistically different. Some rhythms were also judged to be more well-formed than

others ($F(6, 293) = 4.20, p = .0005$); R4 was judged the most well-formed, and R7 the least well-formed; the others were not significantly different. The familiarity and well-formedness ratings for rhythms correlated significantly with each other ($F(1, 299) = 106.89, p < .0001$). Modes were also significantly different in terms of familiarity ($F(6, 293) = 9.81, p < .0001$); M1 was rated as the most familiar mode; the others were not different statistically. The same pattern held for well-formedness ratings which also differed significantly $F(6, 293) = 15.27, p < .0001$. Well-formedness judgments correlated significantly with familiarity ($F(1, 299) = 234.68, p < .0001$). These results are similar to those in Experiments 1 and 2.

Combinations of Rhythms and Modes

To test whether listeners prefer large pitch intervals to coincide with long tone durations, we coded the surface-level match in three ways. We first coded all isomorphic combinations (rhythms and modes that share the same structure in Table 2.1) as 1 and all non-isomorphic combinations (rhythms and modes that do not share the same structure in Table 2.1) as 0. The second surface-level coding was the number of times that pitch interval and tone duration matched. For example, as Table 2.1 shows, R1 and M4 matched for the first two intervals (2 2 and T T, respectively) and the last three intervals (2 2 1 and T T S, respectively), making a total of 5 matches; R1-M4 was coded as 5. The third surface-level coding was based on the number of matches prior to the first mismatch. For example, as Table 2.1 shows, R1 and M4 share the same structure (2 2 and T T, respectively) at the beginning of the sequence, and diverged at the third interval/duration (1 and T, respectively), making a total of 2 matches prior to the first mismatch; R1-M4 was coded as 2.

The correlation between the surface-level match and the fit judgments was not significant for the first two types of surface-level coding (the first type: $F(1, 2399) = 1.21, p = .2715$; the

second type: $F(1, 2399) = .1803, p = .6711$), but it was significant for the third type, the number of events that match before they diverge ($F(1, 2399) = 13.01, p = .0003$). The high ratings for both probe accents and probe tones in the initial serial positions suggest a focus on early events in the sequence might be driving this effect. The correlation was no longer significant when these cases were excluded.

To test whether listeners preferred tonally stable tones, as measured in Experiment 2, to coincide with temporal positions where accents were preferred, as measured in Experiment 1, we correlated the probe tone ratings from Experiment 2 with the probe accent ratings from Experiment 1 for each pair of modes and rhythms. We call these correlations the *hierarchical-level match*. For any particular pair, the correlation would be strong if the tonally strong events occurred when rhythmic accents were preferred. The pair R1 and M4 (Figures 2.7 and 3.2) is a case in which tonally strong events occur where accents are preferred. In contrast, R3 and M2 have dissimilar ratings, so their combination would place tonally strong events where accents are not preferred. The hierarchical-level match had a strong effect on the fit judgments for individual participants ($F(1, 2350) = 203.04, p < .0001$), shown in Figure 4.1. This means that the measure derived from Experiments 1 and 2 (the degree to which their hierarchies agreed) predicted well how the rhythm was judged to fit the scale.

In addition, the fit judgments correlated with the familiarity of the rhythm ($F(1, 1926) = 47.76, p < .0001$), the familiarity of the mode ($F(1, 2161) = 108.86, p < .0001$), the well-formedness judgment of the rhythm ($F(1, 2184) = 67.75, p < .0001$), and the well-formedness judgment of the mode ($F(1, 2318) = 158.87, p < .0001$). When all these and the hierarchical-level match were entered into a mixed-model multiple regression together with the fit judgments, the

hierarchical-level match was the strongest predictor of the fit judgments ($F(1, 2410) = 132.96, p < .0001$).

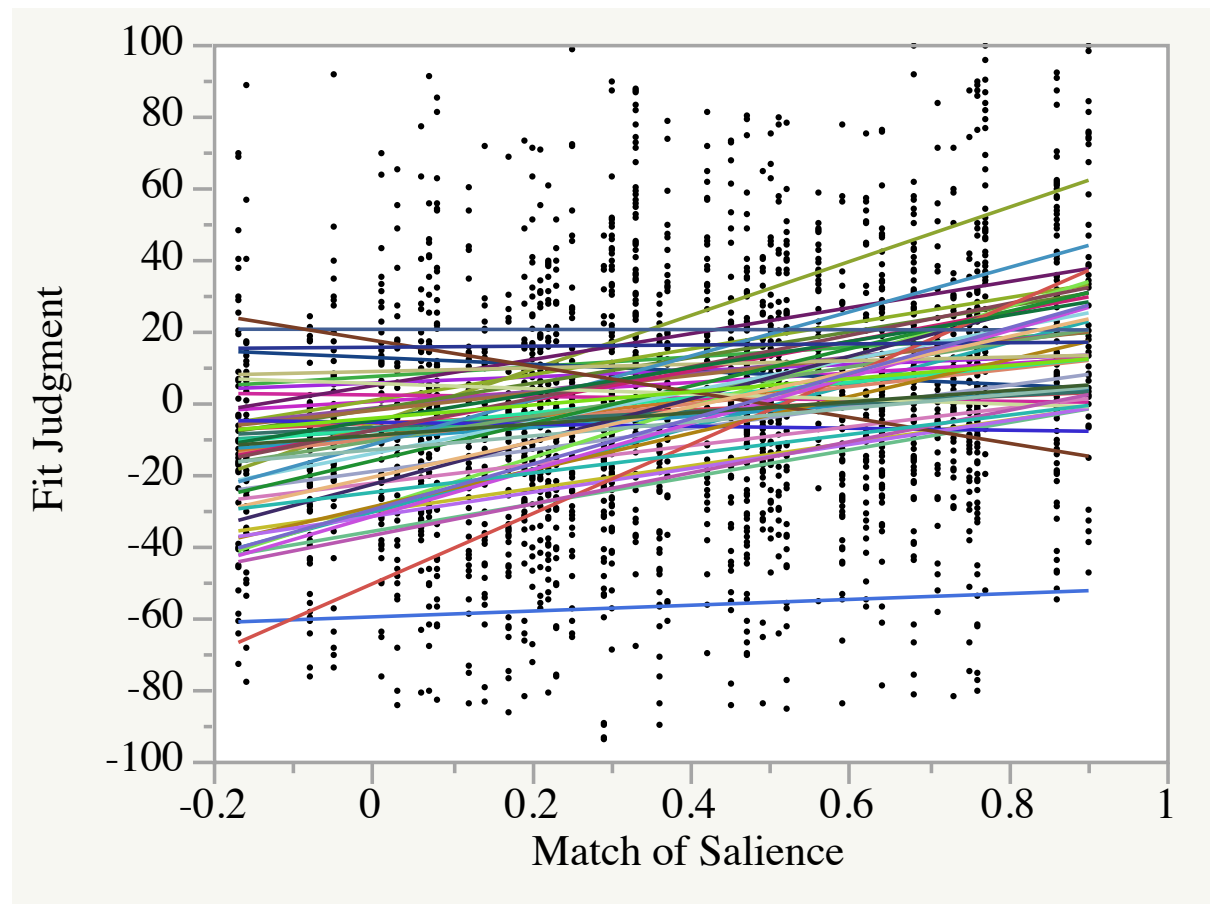


Figure 4.1. Correlation between the fit judgment from Experiment 3 (y-axis) and the hierarchical-level match that is derived from the probe accent ratings in Experiment 1 and the probe tone ratings in Experiments 2 (x-axis). The fit lines indicate the correlation trends for individual participants.

Discussion

Listeners preferred tonally strong events (tones that were judged as fitting well with the mode contexts) to occur where temporal accents (accents that were judged as fitting well with the

rhythm contexts) are preferred. Thus, the hierarchical-level match influenced the fit judgments. The results suggest an interaction at a hierarchical level between pitch and time. This was true even when familiarity and well-formedness of the rhythms and modes were taken into account. The findings did not support the idea that surface-level matches influenced the fit judgments; combination of rhythms and modes where large intervals corresponded with longer durations were not preferred.

This experiment provides another example of perceptual interactions between the pitch and time dimensions, rather than independent processing of the two dimensions. The result is in accord with explanations for the divergent findings in the literature (Prince, 2011; Prince, Schmuckler, & Thompson, 2009). The task of judging how well the rhythm fits the mode would rely on late processes because it focuses on the relationship between the two patterns, requiring the integration of the two dimensions. The task also emphasizes global processing, as the entire rhythmic sequence is to be compared with the entire mode. Because the rhythms and modes are dominant in their respective musical cultures, it could be argued that they are both cognitively coherent patterns. In addition, the two dimensions are matched both on the surface, where they contain the same number of elements with two types of intervals (2 tone durations and 2 interval sizes), and on deeper levels, where they share many structural properties. Given this, the two dimensions would likely be fairly well matched in salience. All these factors help understand why the results in the current experiment depended on the interaction on a hierarchical level.

DISCUSSION FOR EXPERIMENTS 1, 2, AND 3

The isomorphism between the standard rhythmic pattern in African music and the diatonic pattern in Western music has been a source of considerable interest and has stimulated a substantial body of music theory. Pressing (1983) argued that beyond the theoretical properties

that these two patterns share (e.g., the isomorphic surface structure of 2 2 1 2 2 2 1), they are likely to derive from shared cognitive principles. In contrast, London (2002) argued against a deep parallel between the standard rhythms and the diatonic modes. Through a set of three experiments, we found support for both viewpoints at different levels: hierarchies within the rhythms and modes, distances between the rhythms and between the modes, and the hierarchical-level match between rhythms and modes.

We turn first to experimental results that suggest different cognitive principles apply to pitch and time. Listeners tended to break the standard rhythmic pattern into three subgroups, suggesting a preference for duple grouping. Accents were preferred on the first event, on a long event following a short event, but not on the penultimate tone. In addition, the syncopation-shifted 3/2 metrical hierarchy proposed by Temperley (1999, 2000) had a consistent perceptual influence on the seven rotations of the standard pattern, especially when it does not occur until the second half of the sequence after a strong metrical context has been formed. The coexisting preferences for subgrouping and the principle of the syncopation-shifted 3/2 metrical hierarchy provide evidence for multiple cognitive principles working together, that is, multistability (Pressing, 1983).

In contrast, the probe tone ratings for the modes were dominated by relationships relative to the tone C, which began and ended sequences in both the ascending and descending contexts. Probe tones close to the initial C in the context, or close to it on either the chroma circle or the circle of fifths, were generally rated more highly; these have been shown to contribute to judgments of inter-tone distance (Krumhansl, 1990). The probe tone ratings also resembled to some extent the tonal hierarchies of C major and C minor (Krumhansl & Kessler, 1982), and the major keys corresponding to the number of sharps and flats (especially when M6 is minor).

These different cognitive principles found for rhythmic and tonal hierarchies support London's (2002) view of non-isomorphisms between pitch and time in perception.

However, the rhythms and the modes were similar in how they were judged in terms of familiarity and well-formedness. Mode 1 and Rhythm 4 were rated as the most familiar and well-formed. Mode 1 is the major scale, namely, the most prominent scale in Western music. Rhythm 4 starts with the longest run of long durations, establishing a stable and regular pattern at the onset, which is consistent with the run principle and typical of Western rhythms. At the other extreme, Mode 7 and Rhythm 7 were rated as the least familiar and well-formed. Mode 7 has the greatest number of accidentals among the modes. Rhythm 7 violates the run principle by starting with a short duration followed by a short run of long durations, beginning with the most irregular pattern. In addition, the modes and rhythms were similar in that the probe ratings were not influenced by pitch interval or temporal interval.

Another perceptual isomorphism was found at the level of inter-rhythm and inter-mode distances. The perceptual distance measure derived from the probe data corresponded best with the swap distance measure for both rhythms and modes, which arranges the rhythms and the modes in this order: R4/M4, R1/M1, R5/M5, R2/M2, R6/M6, R3/M3, R7/M7. This ordering corresponds to the relative number of sharps and flats. It also follows the emotion connotations of the modes from happy to sad found by Temperley and Tan (2013) except that M1 was judged happiest, perhaps because of familiarity. Finally, swap distance, a computational model of rhythmic distance, was found to correspond to the arrangement of the diatonic modes based on the species of the 4th and the species of the 5th, coming from a very different theoretical tradition.

A similar isomorphism between time and pitch was found in a recent study of polyrhythms and pitch intervals (Razdan & Patel, 2016). They translated the frequency ratios of the intervals from minor second to octave (plus major, minor, augmented, and diminished chords) into polyrhythms. The ratings of the polyrhythms for pleasantness correlated strongly with pleasantness ratings for the corresponding pitch intervals.

Finally, the fit judgments revealed that the pitch-time parallel extends across dimensions. When the modes were paired with the rhythms, listeners preferred combinations in which rhythmic and tonal hierarchies coincided, that is, where tonally strong events occurred at point in the rhythms where accents were preferred. Listeners did not prefer combinations in which larger pitch intervals were separated by longer durations. Thus, fit judgments depended on the hierarchical-level match, not the surface-level match. That pitch and time interact in this way is consistent with a number of proposals about what task and stimulus characteristics favor perceptual interactions between dimensions rather than independent processing: late processing, global judgments, coherent patterns, abstract rather than surface properties, and balanced salience of the dimensions.

In addition to the musical features that were tested, at least three other theoretical properties are shared by the standard and diatonic patterns. The first characteristic is *maximally even set*. The seven positions in each standard rhythm and the seven pitches in each diatonic mode both form a set that is the closest to being spread evenly around the circle (Clough & Douthett, 1991; Toussaint, 2013). In other words, the 7 dark circles in Figures 2.1 and 2.2 cannot be more evenly distributed around the 12 clock positions or tone identities, and thus form a maximally even set. The second characteristic is that the standard and diatonic patterns are both *asymmetrical*. Each tone has a unique set of relations from every other tone (Browne, 1981; Rahn, 1996; Temperley,

2000). As a result, for the scales, the asymmetrical distribution of the semitones facilitates position-finding based on only two or three tones. The function of position-finding might also be applicable to the rhythms. The third characteristic is that the standard pattern is a *deep rhythm* (Toussaint, 2013) and the diatonic pattern is a *deep scale* (Gamer, 1967; Johnson, 2003; Toussaint, 2005). For a pattern to be *deep*, the distance between every pair of onsets or pitches has to occur a unique number of times. All three theoretical properties suggest there may be other aspects of the isomorphism between standard and diatonic patterns that might be tested in future perceptual studies.

CHAPTER 5

ISOMORPHISM IN THE BELL AND PENTATONIC PATTERNS

The series of three experiments reported in Chapters 6, 7, and 8 was initially motivated by the findings in Experiments 1, 2, and 3. Because these experiments uncovered cognitive principles underlying the perception of pitch and time and their interaction in the standard and diatonic patterns, we wanted to find out the extent to which these principles can be generalized to other patterns with structural isomorphism. The bell rhythmic pattern and the pentatonic scale pattern share the surface structure of 2 2 3 2 3. This structure describes both the temporal intervals between every two consecutive events in the bell pattern (Figure 5.1), and the pitch intervals between every two consecutive tones in the pentatonic pattern (Figure 5.2).

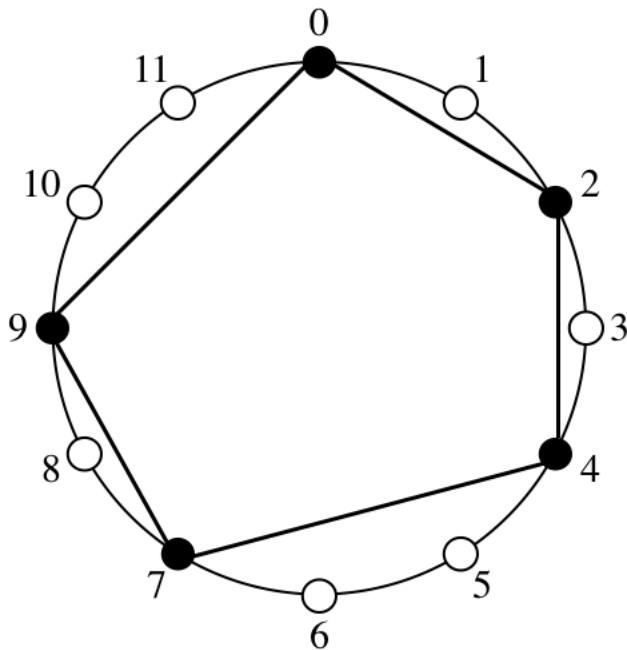


Figure 5.1. Theoretical structure of the bell pattern in clock position. The black circles mark temporal events that sounded in the bell pattern. Starting from clock position 0 going clockwise, the lines connecting consecutive black circles form durations of 2 2 3 2 3 (redrawn from Toussaint, 2013, Figure 14.7).

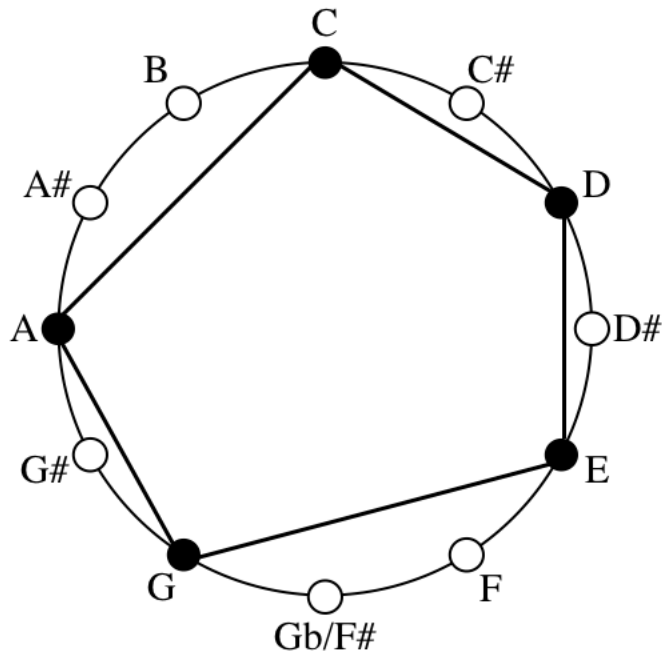


Figure 5.2. Theoretical structure of the pentatonic pattern in tone identity. The black circles mark pitches that sounded in the pentatonic pattern. Starting from the pitch C going clockwise, the lines connecting consecutive black circles form semitone intervals of 2 2 3 2 3 (redrawn from Toussaint, 2013, Figure 14.7, with modifications).

The bell pattern is the timeline that the standard pattern derived from (Pressing, 1983), and has been viewed as a typological feature of African music generally (Jones, 1959, pp.213). The bell and pentatonic patterns share the same characteristics that the standard and diatonic patterns share. Both patterns are rotational, where they can begin with each successive duration or interval and form five unique rhythms or modes, respectively. For example, Rhythm 1 has the structure of 2 2 3 2 3, Rhythm 2 starts with the second duration in Rhythm 2, generating the rotation 2 3 2 3 2, and so on (see Figure 5.3). Pressing (1983, Table II) and Toussaint (2013, p.77-78) reported that all five rotations of rhythms are found in world music, including African,

Central America, and the Balkan peninsula. Similarly, the five rotations of the scale are also all unique (see Figure 5.4). These pentatonic modes are ubiquitous and can be found in cultural traditions such as Chinese, Balkan, Scottish, Yoruba, and Western, and genres such as jazz, blues, and folk. Thus, perceptual parallels and interactions between pitch and time in the ubiquitous bell and pentatonic patterns not only generalizes existing results to another set of isomorphic patterns but also generalizes the results to broader geographical regions and cultural traditions.

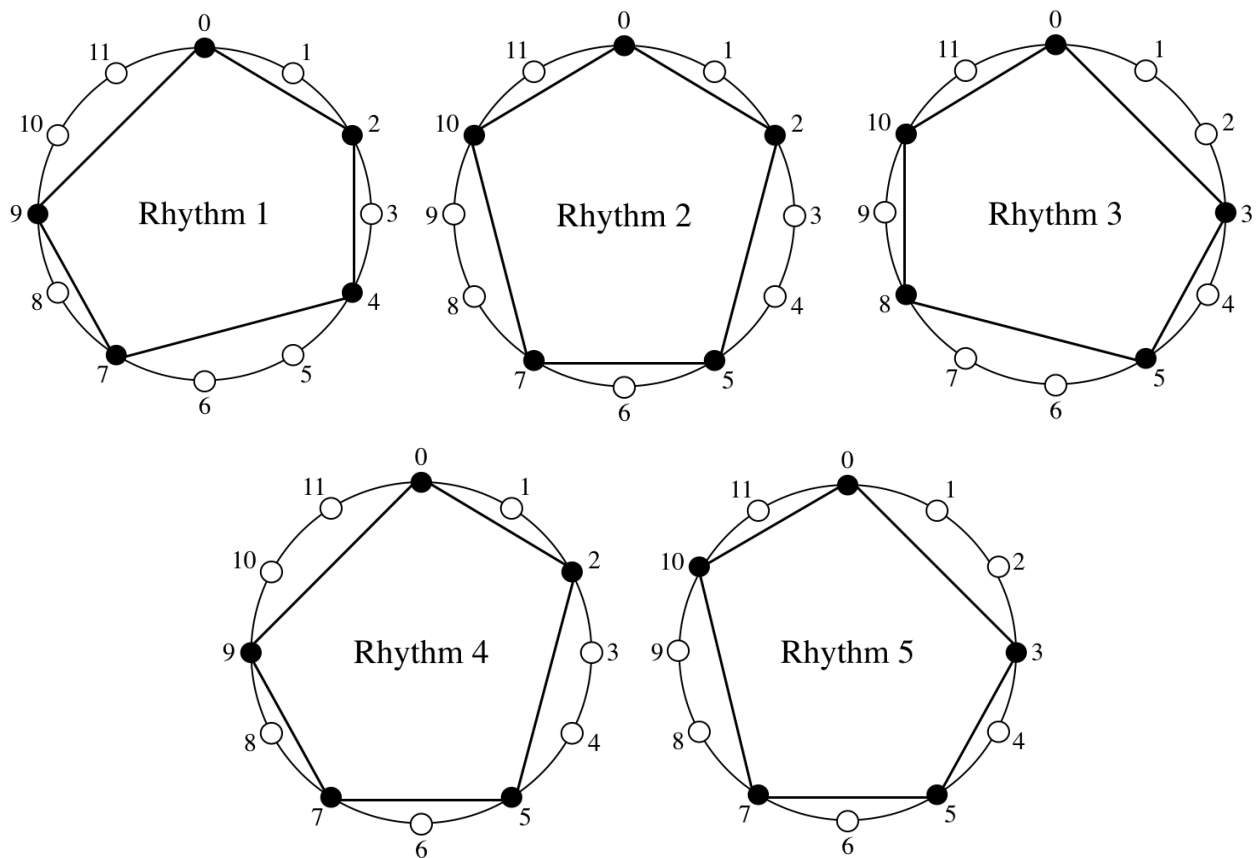


Figure 5.3. The five unique rhythms of the bell pattern in clock position. Each of the five rotations starts with a successive duration. For example, Rhythm 2 starts with the second duration in Rhythm 1 (redrawn from Toussaint, 2013, Figure 14.7, with modifications).

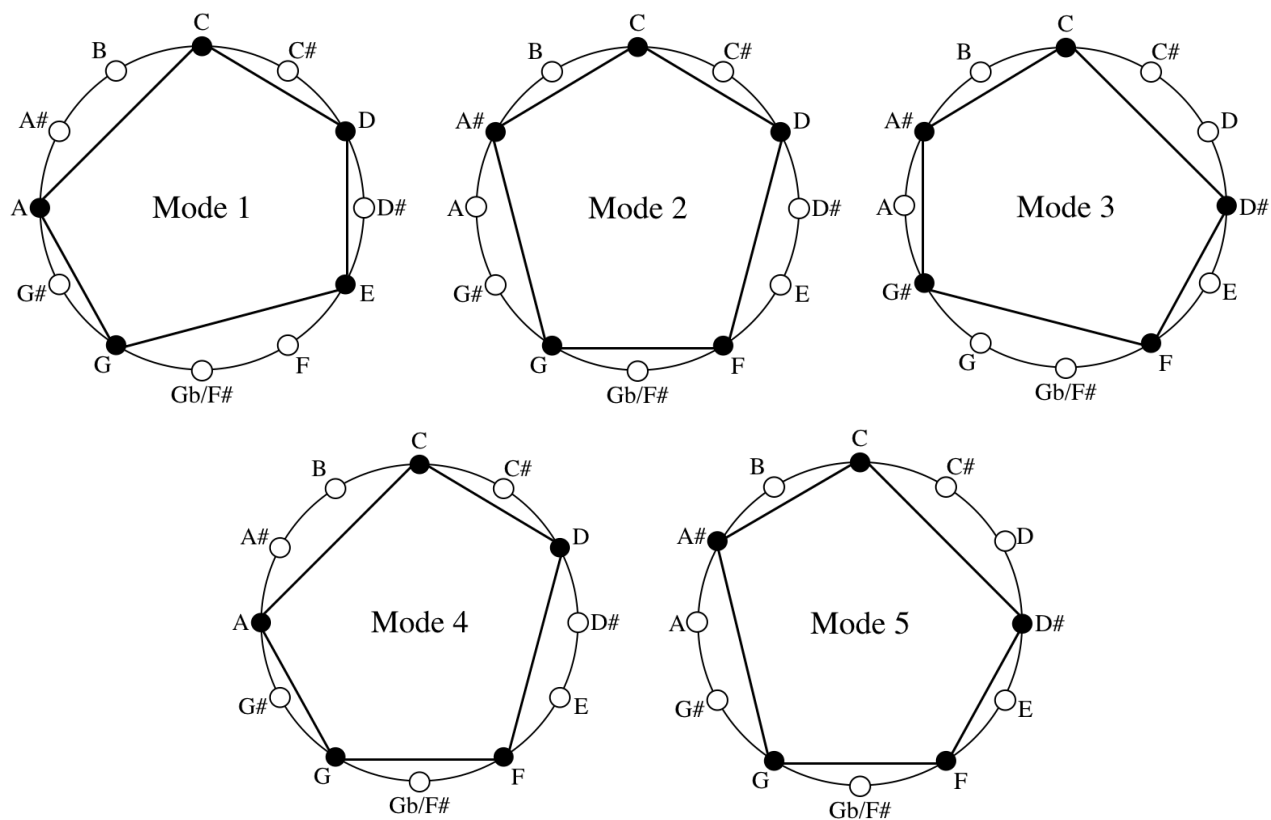


Figure 5.4. The five unique modes of the pentatonic pattern in tone identity. Each of the five rotations starts with a successive interval. For example, Mode 3 starts with the third interval in Mode 1 (redrawn from Toussaint, 2013, Figure 14.7, with modifications).

Three experiments were conducted to find out to what extent the bell rhythms and the pentatonic modes are governed by the same cognitive principles. We asked the same two questions as for the standard rhythms and the diatonic modes. Would the five isomorphic combinations be preferred over other combinations in which the surface structures are not isomorphic? Would the listeners prefer combinations in which tonally strong events occur at points of rhythmic stress? Two experiments were conducted to examine the latter possibility by determining the hierarchies of events in the rhythms and the hierarchies of tones in the modes.

The same music theoretical proposals that were tested in Experiments 1 and 2 were compared with the perceived hierarchical organizations in Experiments 4 and 5. Also tested were predictions about the relationships between the rhythms and between the modes.

Experiment 4a&b uses the probe accent method to find out where accents are preferred in each bell rhythm, which will be detailed shortly in this chapter. Experiment 5a&b uses the probe tone method to find out the tonal stability of each tone in each pentatonic mode, which will be detailed in Chapter 6. Finally, Experiment 6 collects the fit judgments for all possible combinations of rhythms and modes. As in Experiment 3, we tested both the influence of surface isomorphisms between tone duration and interval size, and the influence of hierarchical match between rhythmic hierarchy (as measured in Experiment 4) and tonal hierarchies (as measured in Experiment 5). Together with Experiment 6, Chapter 7 will discuss the results of Experiments 4, 5, and 6. The discussion includes differences and similarities in the cognitive principles underlying the perception of pitch and time, and future directions.

EXPERIMENT 4A&B

As noted above, the bell pattern, from which the standard pattern is derived, can be viewed as a rhythmic sequence with the temporal intervals of 2 2 3 2 3. The bell pattern originated in the same geographical region as the standard pattern (Agawu, 2006; Kubik, 1999; Pressing, 1983; Rahn, 1987). The bell pattern is a timeline, which is played in a repetitive and continuous fashion that serves as a temporal reference (Agawu, 2006; Pressing, 1983; Rahn, 1987). It possesses all characteristics of the standard rhythmic pattern that were mentioned in Chapter 2.

The bell rhythmic pattern is *rotational*, where any of the five events can serve as the first beat of the rhythm, generating five unique patterns (Pressing, 1983). The starting point varies across cultures and geographical regions. Pressing (1983, Table II, patterns 3, 4, 5, and 6) summarized

regions in West Africa where four out of the five rhythms occur, and Toussaint (2013, p.77-78) details countries in Africa, Central America, and the Balkan peninsula where each of the five rhythms occur. The first three columns in Table 5.1 show the structure and the rhythmic notation of the five rhythms, where Rhythm 1, Rhythm 2, ..., Rhythm 5 are abbreviated as R1, R2, ..., R5. Figure 5.3 illustrates the same rhythms in terms of clock position. Based on the three properties that characterized the standard pattern in Experiment 1 (Kubik, 1999), the bell pattern also has 12 as the cycle number but with only 5 events in each cycle, and the asymmetric distribution generates two adjacent sub-patterns with unequal durations (i.e., for the 2 2 3 2 3 rotation, the first sub-pattern has a duration of 7, two 2's and one 3, and the second sub-pattern has a duration of 5, one 2 and one 3).

Because the bell pattern is often briefly mentioned as a related rhythm of the standard pattern, it has received less attention from scholars compared to the standard pattern. No empirical research that we are aware of has addressed how the bell rhythms are perceptually organized, either in terms of grouping or hierarchical patterns of stress. However, the theoretical proposal of tone duration for the standard pattern can be adapted to the bell pattern (Agawu, 2006). In the bell rhythms, the long tone to short tone ratio for the bell rhythms is 3:2 as opposed to 2:1 for the standard rhythms. However, because the proposals about the three metrical hierarchies were specific to the standard rhythms (Temperley, 2000), not the bell rhythms, they will not be tested here. Accordingly, because the syncopation shift builds on metrical hierarchies, it will not be discussed for the bell rhythms.

In addition, the proposal of swap distance about how the different rotations relate to each other can be applied to the bell rhythms (Toussaint, 2013). Swap distance is the minimum number of swaps between adjacent pulse units necessary to transform one rhythm into another.

Table 5.1. *Structure and stimuli of the bell and pentatonic patterns.*

Bell pattern			Pentatonic pattern		
Structure	Rhythm	Temporal arrangement	Structure	Mode	Pitch arrangement
2 2 3 2 3	Rhythm 1 (R1)	♩ ♩ ♩. ♩ ♩. (♩)	T T 3S T 3S	Mode 1 (M1)	C D E G A (C)
2 3 2 3 2	Rhythm 2 (R2)	♩ ♩. ♩ ♩. ♩ (♩)	T 3S T 3S T	Mode 2 (M2)	C D F G A [#] (C)
3 2 3 2 2	Rhythm 3 (R3)	♩. ♩ ♩. ♩ ♩ (♩.)	3S T 3S T T	Mode 3 (M3)	C D [#] F G [#] A [#] (C)
2 3 2 2 3	Rhythm 4 (R4)	♩ ♩. ♩ ♩ ♩. (♩)	T 3S T T 3S	Mode 4 (M4)	C D F G A (C)
3 2 2 3 2	Rhythm 5 (R5)	♩. ♩ ♩ ♩. ♩ (♩.)	3S T T 3S T	Mode 5 (M5)	C D [#] F G A [#] (C)

Table 5.2 shows the swapping process between rhythms, arranged in the linear order of R1, R4, R2, R5, and R3, with the swapped positions highlighted. Table 5.3 shows the swap distance between every two rhythms. Exchanging “halves” was not tested for the bell rhythms because it was only proposed for the standard rhythms (Agawu, 2006).

Objectives of the Study

Experiment 4 was designed to explore the perceptual organization of the bell pattern, in particular, hierarchies of accents in the rhythms. As in Experiment 1, we used the probe accent technique to probe each of the first five temporal positions in each of the bell rhythms. The listeners’ task was to rate how well the probe accent fits into the rhythm. The resulting data will be considered in light of the various proposals tested in Experiment 1 for how events within the rhythms are organized and how the five rhythms are related to one another.

Method for Experiment 4a

Participants

Forty-eight students at Cornell University (15 males and 33 females, age 18-40) participated in Experiment 4a for course credit or cash. On average, they have played music for 14.38 years (including music lessons; summed over all instruments and voice; range, 0-38; four had absolute pitch, and three had no musical training). The study was approved by Cornell University’s Institutional Review Board.

Apparatus and Stimulus Materials

We constructed five rhythms based on the bell pattern by shifting the starting tone duration (see Figure 5.3 and Table 5.1). All bell rhythms were composed similarly to the standard rhythms in Experiment 1. The only exceptions are that the bell rhythms consisted of 6 monotonic tones on C4, forming 5 time intervals, with the probe accent on one of the first five events in

Table 5.2. *The swapping process for the bell rhythms and pentatonic modes.*

Bell rhythms in clock position													
Rhythm	0	1	2	3	4	5	6	7	8	9	10	11	
R1	1	0	1	0	1	0	0	1	0	1	0	0	M1(0#, 0b)
R4	1	0	1	0	0	1	0	1	0	1	0	0	M4(0#, 0b)
R2	1	0	1	0	0	1	0	1	0	0	1	0	M2(1#)
R5	1	0	0	1	0	1	0	1	0	0	1	0	M5(2#)
R3	1	0	0	1	0	1	0	0	1	0	1	0	M3(3#)
	C	Db	D	Eb	E	F	F#/Gb	G	Ab	A	Bb	B	Mode (no. of accidentals)
Pentatonic modes in tone identity													

Table 5.3. *Bell rhythms and pentatonic modes arranged according to swap distance.*

Rhythm	R1	R4	R2	R5	R3	
R1	0	1	2	3	4	M1(0#, 0b)
R4	1	0	1	2	3	M4(0#, 0b)
R5	2	1	0	1	2	M2(1#)
R5	3	2	1	0	1	M5(2#)
R3	4	3	2	1	0	M3(3#)
	M1 (0#, 0b)	M4 (0#, 0b)	M2(1#)	M5(2#)	M3(3#)	Mode (no. of accidentals)

each rhythm. All rhythms were at the tempo of 140 quarter-note per minute (each quarter note lasts approximately 440 ms, each trial lasts approximately 7.48 s). Also constructed were 5 unaccented *neutral* rhythms.

Procedure

As in Experiment 1, participants rated how well the probe accent fits into the rhythm after listening to each of the 25 probe accent trials, which were blocked by rhythm type (i.e., R1, R2, etc.). Both the 5 blocks and the 5 probe accent trials within each block were presented in a randomized order. Next, participants listened to each of the 5 neutral (unaccented) rhythms in a randomized order and rated their familiarity and well-formedness as in Experiment 1. Experiment 4a lasted less than 15 minutes. Participants were randomly assigned so that half of them completed Experiment 4a first and the other half completed Experiment 5a first, which will be described later. At the end of the experimental session, participants filled out a demographics questionnaire. Together with Experiment 5a, the experiment lasted approximately 30 minutes.

Method for Experiment 4b

Participants

Forty-four students at Cornell University (13 males and 31 females, age 18-30) participated in Experiment 4b for course credit or cash. On average, they have played music for 10.86 years (including music lessons; summed over all instruments and voice; range, 1-28.5; four had absolute pitch). The study was approved by Cornell University's Institutional Review Board.

Apparatus and Stimulus Materials

In addition to the same bell rhythms used in Experiment 4a, we constructed all rhythms on C5. This was to match the combined ascending and descending scales within each trial in

Experiment 5b, which will be described later. All rhythms were at the tempo of 180 quarter-note per minute (each quarter note lasts approximately 345 ms, each trial lasts approximately 5.87 s).

Procedure

The procedure was identical to Experiment 4a, except each rhythm was played and rated twice, once on C4 and once on C5. Together with Experiment 5b, the experiment lasted approximately 30 minutes.

Results

All continuous rating scales were coded from -100 to 100. Ratings from the three forms of rhythms all significantly correlated with each other (4a & 4b_C4: $r(23) = .67, p = .0002$; 4a & 4b_C5: $r(23) = .76, p < .0001$; 4b_C4 & 4b_C5: $r(23) = .90, p < .0001$). Thus, all data analysis used the average rating across the three forms of rhythms. Ratings from only 29% of participants correlated significantly with the average probe accent ratings, suggesting there may be some individual differences in the pattern of responses. To investigate this, a clustering analysis (Ward method) was done on the individual data and three main clusters were found. These were compared to demographic factors, including possession of absolute pitch, years of training on musical instruments or voice, whether they have taken a music theory class, whether they have been in a class where African rhythms were mentioned, their familiarity with music or dance related to the standard pattern (Afro-Caribbean music, Jazz, Latin Dance, and African Dance). However, only one of these was different between the groups (familiarity with African Dance: $F(2, 129) = 4.64, p = .01$) and one was marginally significant between the groups (familiarity with Afro-Caribbean music: $F(2, 129) = 2.87, p = .06$).

A related question is whether some rhythms were more familiar than others prior to the experiment. To test this, we conducted a preliminary analysis of the familiarity ratings on the

five neutral (unaccented) rhythms and found significant differences between rhythms ($F(4, 586) = 6.70, p < .0001$). A post-hoc test (Tukey HSD) showed that R1 was rated as the most familiar and R3 was rated as the least familiar; the remaining rhythms were not statistically different from each other. As can be seen in Table 5.1, R1 is the rhythm that starts with the longest string of short durations, establishing a stable pulse. In contrast, R3 begins by alternating between long and short durations. A different ordering was found for the well-formedness ratings, where R2 was rated as the most well-formed and R5 was rated as the least well-formed; the remaining rhythms were not statistically different from each other. R2 is the only rhythm that has a symmetrical structure. R5 starts with a single long duration, followed by a short duration, and then alternates twice more between short and long durations at the end. Despite these differences, R1 and R2 received higher ratings than R3, R4, and R5 in both ratings. The two ratings correlated significantly with each other ($F(1, 678) = 172.43, p < .0001$).

Hierarchies within the rhythms

Figure 5.5 shows the probe accent ratings for each of the five rhythms plotted according to serial position and labeled by clock position. For all the rhythms, there is an overall trend of decreasing ratings from earlier events to later events. The exceptions are a few peaks, such as the third and the fifth events in R1, the second and the fourth events in R2, the third event in R3, the last event in R4, and the fourth event in R5. All these peaks take place on long tone durations (“3”), as the values on the x-axis in Figure 5.5 indicate, also shown in Figure 5.3. In the statistical analysis, serial position was coded as a linear variable, tone duration and rhythm were coded as categorical variables, and participants as random. A mixed-model multiple regression on the probe accent data with tone duration, serial position, and rhythm as factors was significant ($F(95, 3302) = 15.00, p < .0001$), and all three factors were significant. Listeners preferred

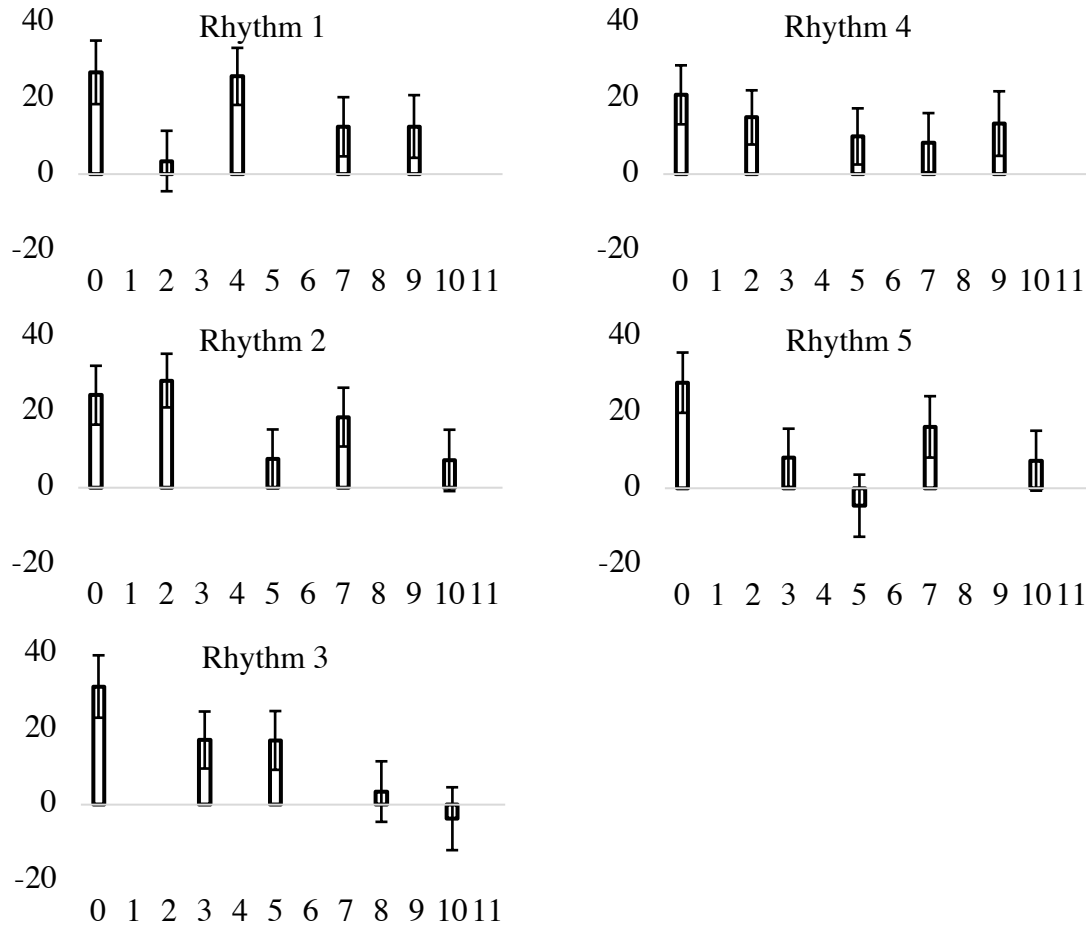


Figure 5.5. Probe accent ratings of each rhythm (y-axis; full range [-100, 100]) in the bell pattern labelled by clock position (x-axis). Error bars indicate 95% confidence interval.

accents on long durations ($F(1, 3302) = 57.96, p < .0001$), and on events earlier in the rhythm ($F(1, 3302) = 69.43, p < .0001$). Rhythm was also significant ($F(4, 3302) = 2.72, p = .03$) with R2 having the highest accent ratings and R5 having the lowest ratings. These two rhythms also happened to receive the highest and the lowest well-formedness ratings, respectively.

Distances between rhythms

As in Experiment 1, we calculated city-block distance to quantify the perceived distances among the five rhythms. This matrix was then correlated with the theoretical measure of swap distance (see Table 5.3). Swap distance did not significantly account for the variation in city-block distances ($R^2 = .06$, $F(1, 8) = .55$, $p = .4798$). Because swap distance orders the rhythms as R1, R4, R2, R5, and R3, the results suggest that listeners did not perceive the bell rhythms in this arrangement.

Discussion

The probe accent ratings for the five rhythms showed that listeners preferred accents on long durations and early events in the rhythms, and perceived the five bell rhythms based on how well-formed each rhythm is. In addition, the highest familiarity ratings of R1 suggest the preference for beginning a rhythmic sequence with the longest run of identical events (the same tone duration) (Handel, 1974; Preusser, Garner, & Gottwald, 1970).

CHAPTER 6

EXPERIMENT 5A&B

The pentatonic mode pattern is isomorphic to the bell rhythmic pattern in the same way the diatonic mode pattern is isomorphic to the standard rhythmic pattern. The pentatonic pattern possesses most characteristics of the diatonic pattern that were mentioned in Chapter 3. Based on the two properties that characterized the diatonic pattern in Experiment 2 (Agmon, 1989), the pentatonic pattern also has 12 scale steps but with only 5 intervals in each cycle. The pentatonic pattern is also *rotational*, such that the five starting points generate five unique modes (see the last three columns in Table 5.1). Figure 5.4 illustrates the same modes in terms of tone identity (C, C#, D, etc.).

Two out of the three types of features that were tested in Experiment 2 were considered in analyzing the probe tone data for the pentatonic modes. The first type of features is basic features. The first feature is the interval size between each tone and the next tone as measured in semitones. In the case of Experiment 5b where a mode consists of both ascending and descending scales, interval size was coded based on the starting direction of the scale. Specifically, ascending followed by descending was coded based on the ascending scale, descending followed by ascending was coded based on the descending scale. The second feature is primacy, which is how early in the context sequence the probe tone occurred after C. The third feature is distance from C on the chroma circle. The fourth feature is distance from C on the circle of fifths.

The second type of features is tonal features, including all consonance measures, other tonal measures, and the KK C major and C minor profiles that were tested for the diatonic modes in Experiment 2. We did not code the major profile for each mode

corresponding to the number of accidentals. This is because the pentatonic modes only consist of a subset of tones from the major keys, making the mapping conceptually unclear. The third type of feature for the diatonic modes was modal features. They are not considered for the pentatonic modes due to the assumption of 7 tones in each mode instead of 5.

The swap distance (Toussaint, 2013) measure was applied to the pentatonic modes. Table 5.2 illustrates the swapping process and Table 5.3 shows the resulting distance values between every two modes. The ordering of modes corresponds to the number of sharps and flats.

Objectives of the Study

Experiment 5 was designed to determine the tonal hierarchies of the modes using the probe tone technique (Krumhansl & Shepard, 1979), in which a probe tone is played after the mode context. As in Experiment 2, listeners rated how well the probe tone fits into the context using the five pitches in each mode. The resulting data will be analyzed for the effects of two types of features described in Experiment 2: basic and tonal. The data will also be used to test how the five modes are related to one another.

Method for Experiment 5a

Participants

The same participants from Experiment 4a participated in Experiment 5a.

Apparatus and Stimulus Materials

We constructed five modes based on the pentatonic scale by shifting the starting pitch interval (see Figure 5.4 and Table 7.1). All pentatonic modes were composed similarly to the diatonic modes in Experiment 2. The only exceptions are that the pentatonic modes consisted of 6 isochronous tones, each lasting for a quarter note, forming 5 pitch intervals. Only the five tones in each mode were used as probe tones. All modes were at the tempo of 140 quarter-note per

minute (each quarter note lasts approximately 440 ms; each trial lasts approximately 3.96 s).

Also constructed were 5 *neutral* modes, which are non-probed modes in ascending form.

Procedure

As in Experiment 2, participants rated how well the probe tone fits into the scale after listening to each of the 25 probe tone trials, which were blocked by mode type (i.e., M1, M2, etc.). Both the 5 blocks and the 5 probe tone trials within each block were presented in a randomized order. Next, participants listened to each of the 5 neutral modes (no probe tone) in a randomized order and rated their familiarity and well-formedness as in Experiment 2. As mentioned in Experiment 4a, the order of Experiments 5a and 4a was counterbalanced.

Method for Experiment 5b

Participants

The same participants from Experiment 4b participated in Experiment 5b.

Apparatus and Stimulus Materials

We used the same scales from Experiment 5a. However, instead of using the ascending and descending scales as separate stimuli, we constructed all scales in the forms of ascending followed by descending (up+down) and descending followed by ascending (down+up). All modes were at the tempo of 180 quarter-note per minute (each quarter note lasts approximately 345 ms, each trial lasts approximately 5.18 s). Also constructed were 10 *neutral* modes without the probe tone (5 up+down and 5 down+up).

Procedure

The procedure was identical to Experiment 5a. As mentioned in Experiment 4b, the order of Experiments 5b and 4b was counterbalanced.

Results

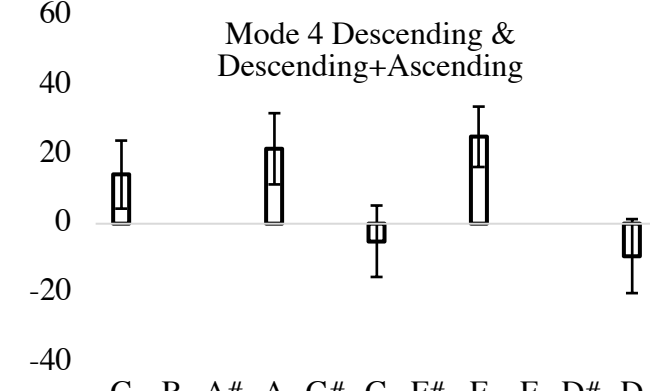
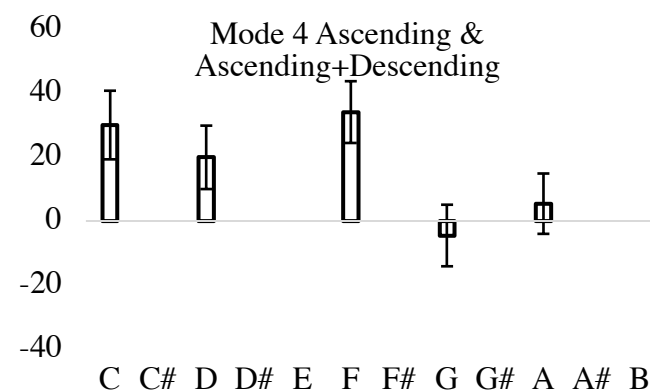
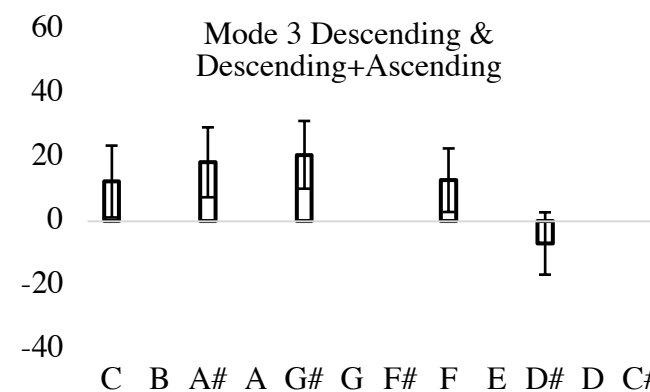
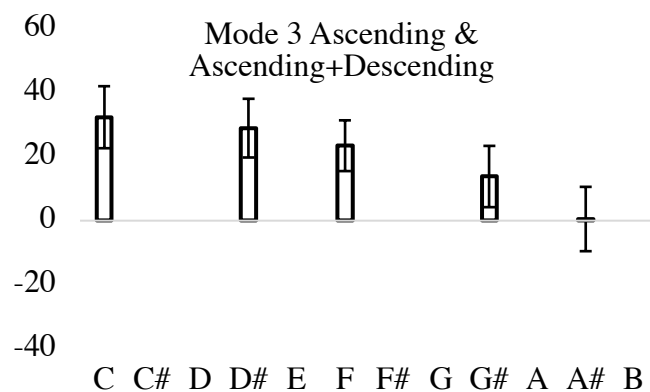
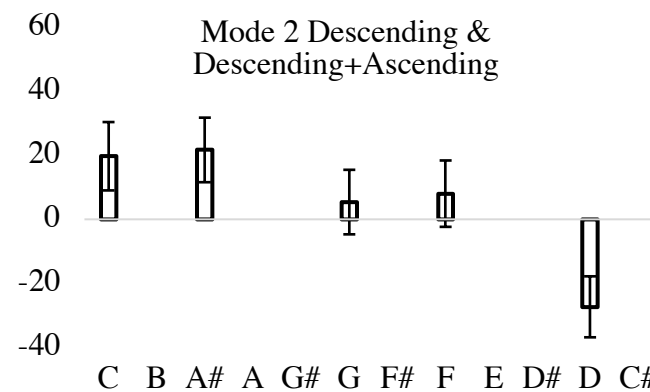
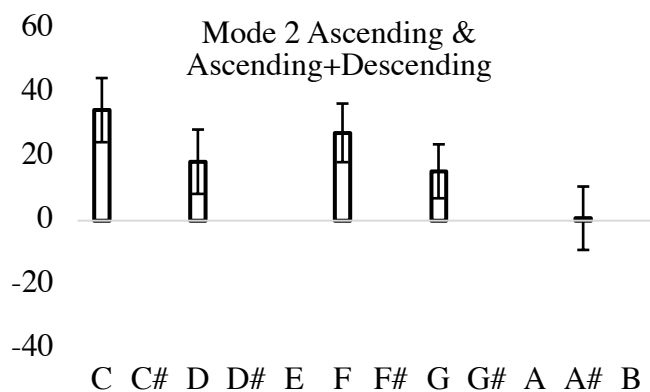
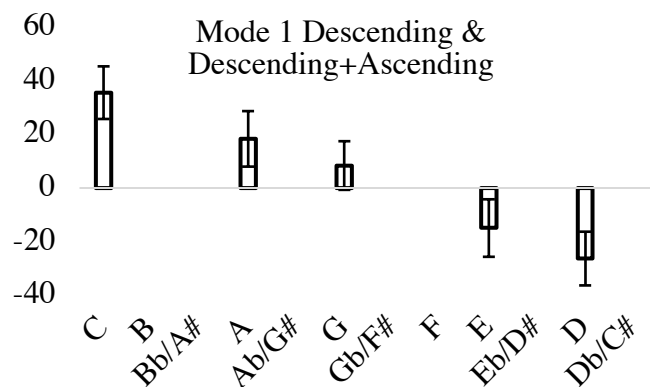
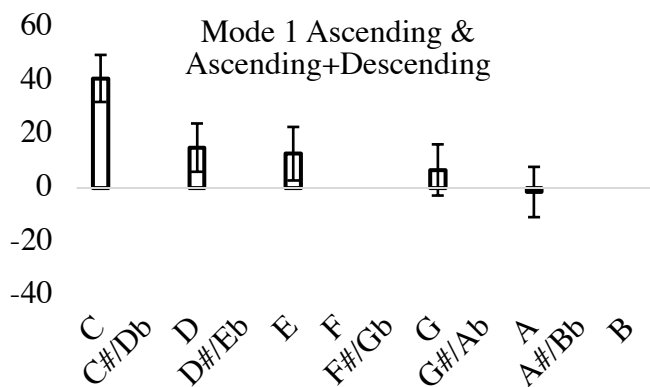
All continuous rating scales were coded from -100 to 100. Ratings from the four forms of modes correlated with each other depend on the direction of the scale (5a_up & 5b_up+down: $r(23) = .73, p < .0001$; 5a_down & 5b_down+up: $r(23) = .70, p < .0001$; 5a_down & 5b_up+down: $r(23) = .44, p < .05$; the rest are not significant). Thus, all data analyses used two sets of probe tone ratings, one set is the average of ascending and ascending followed by descending modes, the other set is the average of descending and descending followed by ascending modes. Ratings from 43% and 47% of participants correlated significantly with the average probe ratings for ascending and ascending followed by descending modes, and for descending followed by ascending modes, respectively. These suggest there may be some individual differences in the pattern of responses.

To investigate this, a clustering analysis (Ward method) was done on the individual data and two main clusters were found for the set of ascending and ascending followed by descending modes, and three main clusters were found for the set of descending and descending followed by ascending modes. These were compared to demographic factors, including whether they had absolute pitch, whether they have ever studied instruments or voice, years of musical training, whether they have taken a music theory class, whether church modes were mentioned in a class they had taken. Of these, no factor differed between the groups for ascending and ascending followed by descending modes. Only one demographic factor was different between the groups for the descending and descending followed by ascending modes (whether or not they were trained in an instrument or voice: $F(2, 89) = 3.62, p = .03$) and one was marginally significant between the groups (training in years: $F(2, 89) = 2.44, p = .09$).

A related question is whether some modes were more familiar than others prior to the experiment. To test this, we conducted a preliminary analysis of the familiarity ratings on the five neutral modes (without probe tone) and found significant differences between modes ($F(4, 586) = 5.03, p = .0005$). A post-hoc test (Tukey HSD) showed that M1 and M5 were rated as the most familiar and M2 and M3 were rated as the least familiar. M1 can be considered as C major and M5 its parallel minor. A different ordering was found for the well-formedness ratings, where M1 and M4 were rated as the most well-formed and M2 was rated as the least well-formed; the remaining modes were not statistically different from each other. M1 and M4 are the only two modes that have no accidentals. Despite these differences between familiarity and well-formedness ratings, M1 received the highest ratings and M2 received the lowest ratings in both ratings. The two ratings correlated significantly with each other ($F(1, 678) = 355.13, p < .0001$).

Hierarchy within a mode

Figure 6.1 shows the probe tone ratings for each of the five modes plotted according to serial position and labeled by tone identity of the probe tone and whether the context starts with ascending or descending scales. There is an overall trend of decreasing ratings from earlier events to later events. Mode 4 shows the most structure in that the ascending and the descending ratings are consistent with each other, where the peaks are on F and A. In the statistical analysis, serial position was coded as a linear variable, interval size and mode were coded as categorical variables, and participants as random. A mixed-model multiple regression on the probe tone data with interval size, serial position, and mode as factors was significant ($F(95, 4503) = 16.38, p < .0001$), and all three factors were significant. Listeners preferred tones with large interval size ($F(1, 4503) = 18.51, p < .0001$), and on earlier tones in the mode context, which is a primacy



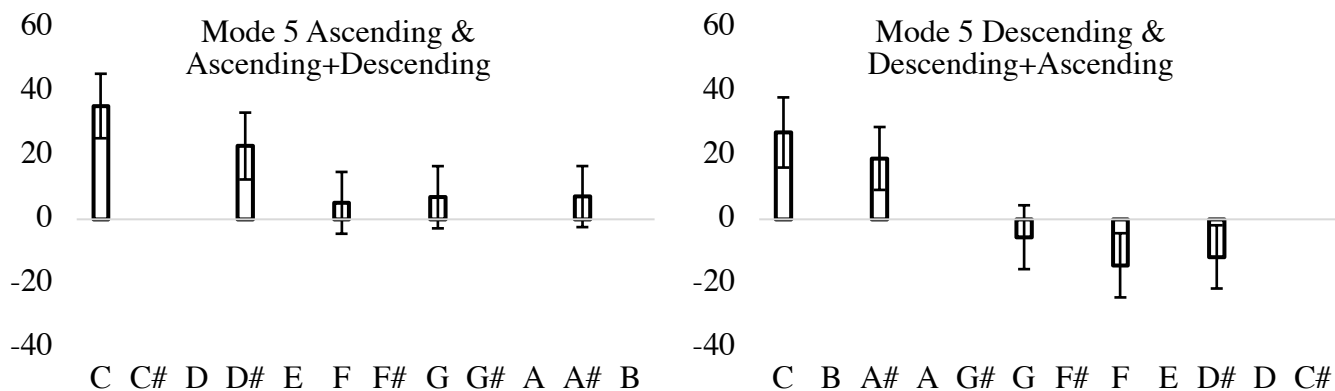


Figure 6.1. Probe tone ratings of each mode (y-axis; full range [-100, 100]) in the pentatonic pattern labelled by tone identity (x-axis), with ascending on the left and descending on the right. Error bars indicate 95% confidence interval.

effect ($F(1, 4503) = 275.74, p < .0001$). Mode was also significant ($F(4, 4503) = 3.48, p = .0076$) with M3 having the highest probe tone rating and M5 and M1 having the lowest ratings.

Basic features. Distance from C on both the chroma circle and the circle of fifths are coded for the pentatonic modes based on the coding in Table 3.1 for the diatonic modes. To find out whether or not these factors contributed uniquely to the probe tone ratings, we conducted a mixed-model multiple regression analysis of the probe tone ratings with each of these two factors together with interval size, serial position, and mode. No feature remained significant in addition to serial position, interval size, and mode (chroma circle: $F(1, 4598) = .08, p = .78$; circle of fifths: $F(1, 4598) = 2.20, p = .14$).

Consonance and tonal features. Each of the consonance measures and tonal measures was tested in a mixed-model multiple regression analysis of the probe tone ratings together with interval size, serial position, and mode. Kameoka and Kuriyagawa (1969, Fig.8) (at $p < .05$), Malmberg's (1918, Table II) Data (at $p < .05$), Malmberg's (1918, Table II) Rank (marg. at p

= .06), Helmholtz's (1885/1954, p. 332) equal-tempered (marg. at $p = .07$), Helmholtz's (1885/1954, p. 332) simple-ratio (at $p < .01$), Hutchinson and Knopoff (1978) (at $p < .0001$), harmonicity (Gill & Purves, 2009) (at $p < .01$), Stolzenburg (2015) (at $p < .05$), and the C minor probe tone profile (Krumhansl & Kessler, 1982, see Figure 1.4) ($p < .05$) all contributed uniquely to the probe tone ratings in addition to serial position and interval size. However, because the high ratings of the tone C might be driving these effects, we conducted the same analyses after removing C. Values from Helmholtz's equal-tempered was no longer marginally significant ($p = .12$), but the other predictors remained significant (at $p < .05$) with Malmberg's rank remaining marginal ($p = .09$). This suggests an effect of consonance and tonal features on the probe tone ratings of the pentatonic modes.

Distances between modes

As in Experiment 4, we calculated city-block distance to quantify the perceived distances among the five modes and applied the swap distance (see Table 5.3) to the modes. Swap distance did not significantly account for the probe tone ratings ($R^2 = .02$, $F(1, 8) = .19$, $p = .6743$). Thus, the pentatonic modes were not perceived in the linear arrangement of M1, M4, M2, M5, and M3.

Discussion

The probe tone ratings for the five modes showed three underlying perceptual influences. The first was how early the probe tone was sounded in the context, the second was the preference for large interval size. The third was consonance measures and tonal hierarchies. These consonance measure had a stronger effect in the perception of the pentatonic pattern compared to that of the diatonic pattern. The effect of tonal hierarchies is weaker here compared to in Experiment 2. These findings suggest that relatively surface-level principles govern listeners' perception of the five pentatonic modes, with some influence of hierarchical-level principles.

CHAPTER 7

EXPERIMENT 6

In exploring how pitch and time combine perceptually, the current experiment is concerned with when listeners find the rhythm as fitting well with the mode. It addresses the same questions as Experiment 3.

Objectives of the Study

Experiment 6 tests whether listeners prefer specific pitches to occur in specific temporal positions using the theoretically isomorphic bell rhythmic pattern and pentatonic mode pattern. Each rhythm was combined with each mode to create a total of 25 (5 x 5) combinations, and listeners judged how well the rhythm fits the scale (fit judgments). Two possible outcomes will be tested, the surface-level match, and the match of salient events and tones from the rhythms and modes, respectively.

Method

Participants

Thirty-nine students at Cornell University (8 males and 31 females, age 18-40) participated in Experiment 6 for course credit. On average, they have played music for 11.6 years (including music lessons; summed over all instruments and voice; range, 2-29; seven had absolute pitch). The study was approved by Cornell university's Institutional Review Board.

Apparatus and Stimulus Materials

We constructed twenty-five rhythm-mode pairs by combining each of the five rhythms in Experiment 4a and each of the five ascending modes in Experiment 5a (see Table 5.1). The range of all sequences was C4 to C5. All modes were at the tempo of 120 quarter-note per minute (each quarter note lasts approximately 500 ms; each trial lasts approximately 3.50 s) and sounded

in piano timbre. In addition, we used the five neutral rhythms from Experiment 4a and the five neutral modes from Experiment 5a as the *neutral* trials for Experiment 3.

Procedure

Participants rated how well the rhythm fits the scale by moving a slider on a continuous scale from extremely bad fit (left end) to extremely good fit (right end). They began with four practice trials. Next, they listened to and rated each of the 25 trials in randomized order. After completing a demographics questionnaire, they then repeated the same task one more time, in a different randomized order. Following the fit judgment, participants completed another demographics questionnaire, and then listened to each of the 10 neutral trials (5 rhythms and 5 modes) in a randomized order and rated them on: 1) how familiar the rhythm/scale is, and 2) how well-formed the rhythm/scale seems, in other words, whether or not the rhythm/scale forms a good pattern. The experiment lasted approximately 30 minutes.

Results

All continuous rating scales were coded from -100 to 100. Ratings from 36% of participants correlated significantly with the average probe ratings, suggesting there may be some individual differences in the pattern of responses. A clustering analysis (Ward method) was done to see if there were identifiable demographic differences between the two clusters that were found. Demographic factors included whether they had absolute pitch, whether they have ever studied instruments or voice, years of musical training, and their familiarity with music or dance related to the standard pattern (Afro-Caribbean music, Jazz, Latin Dance, and African Dance). No factor differed between the groups.

As in Experiments 4 and 5, one question to consider is whether some rhythms or modes were more familiar than others prior to the experiment. We averaged the two repeated measures of fit

judgments because they significantly correlated with each other ($r(23) = .72, p < .0001$). A preliminary analysis of the familiarity ratings on the five neutral (unaccented) rhythms found significant differences between rhythms ($F(4, 152) = 4.18, p = .0031$). A post-hoc test (Tukey HSD) showed that R4 was rated as the most familiar and the remaining rhythms were not statistically different from each other. A similar ordering was found for the well-formedness ratings, where R4 and R2 were rated as the most well-formed and the remaining rhythms were not statistically different from each other. In both familiarity and well-formedness ratings, R4 received the highest ratings and the R1, R3, and R5 received the lowest ratings. The two ratings correlated significantly with each other ($F(1, 193) = 71.83, p < .0001$). In contrast, the five pentatonic modes were not significantly different from each other in their familiarity ratings, and the well-formedness ratings were only marginally significant. The familiarity and well-formedness ratings for both the rhythms and the modes are different from ratings from the probe studies (Experiments 4 and 5). This inconsistency suggests that listeners are uncertain about how to judge the bell rhythms and the pentatonic modes, which likely reflects their overall unfamiliarity with these patterns.

Combinations of Rhythms and Modes

Similarly to the 7/12 pattern in Experiments 1, 2, and 3, we tested the surface-level match using the same measures. The three surface-level coding are the number of exact matches of pitch interval and tone duration, number of times that pitch interval and tone duration matched, and the number of matches prior to the first mismatch. The correlation between the surface-level match and the fit judgments was not significant for any of the three types of surface-level coding.

We also coded the match of salience, which is the match of salient events and tones from the rhythms and modes, respectively. This was calculated by correlating the probe tone ratings from

Experiment 5 with the probe accent ratings from Experiment 4 for each pair of modes and rhythms. Ratings from the three forms of rhythms in Experiments 4a and 4b were averaged because they all significantly correlated with each other. From Experiments 5a and 5b, we only used probe tone ratings of ascending scales and ascending followed by descending scales. This is because these scales are the only scales starting with the ascending direction, which is consistent with scales used in the rhythm-mode combinations in the current experiment. Probe tone ratings of these two types of scales were averaged as they are highly correlated with each other ($r(23) = .73, p < .0001$). Accordingly, correlating the average probe accent ratings and the average ascending probe tone ratings resulted in the match of salience.

Figure 7.1 shows that the match of salience had a strong effect on the fit judgments for individual participants ($F(1, 897) = 8.98, p < .01$). The significant correlation suggests that the measure derived from Experiments 4 and 5 (the degree to which their salience matched) predicted well how the rhythm was judged to fit the scale.

In addition, the fit judgments correlated with the familiarity of the rhythm ($F(1, 935) = 7.84, p < .01$), the familiarity of the mode ($F(1, 935) = 5.48, p < .05$), and the well-formedness judgment of the mode ($F(1, 935) = 27.56, p < .0001$), though the correlation with the well-formedness judgment of the rhythm was not significant ($F(1, 935) = 1.99, p = .1588$). When all these and the match of salience were entered into a mixed-model multiple regression together with the fit judgments, well-formedness judgment of the mode was the strongest predictor of the fit judgments ($F(1, 931) = 23.80, p < .0001$). This suggests a perceptual bias towards pitch, which is consistent with previous evidence on the relative salience of the pitch dimension compared to the time dimension (Prince, 2011; Prince & Pfordresher, 2012; Prince, Schmuckler, & Thompson, 2009).

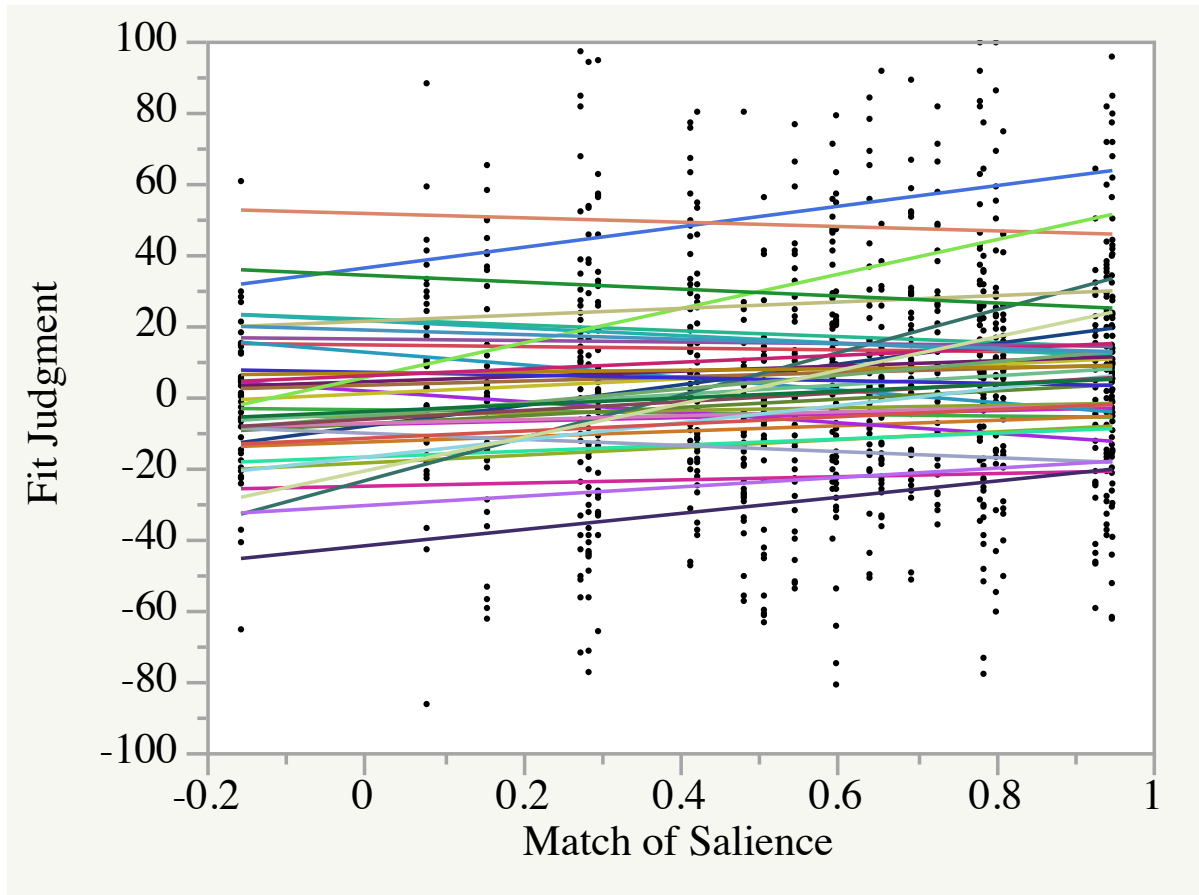


Figure 7.1. Correlation between the fit judgment from Experiment 6 (y-axis) and the match of salience that is derived from the probe accent ratings in Experiment 4 and the probe tone ratings of the ascending and ascending followed by descending modes in Experiments 5 (x-axis). The fit lines indicate the correlation trends for individual participants.

Discussion

Listeners preferred tones that occurred early, formed a large interval size with the following tone, and are consonant with C to occur at early temporal positions and positions with long tone durations. Thus, the match of salient events and tones from the rhythms and modes, respectively, influenced the fit judgments. Interestingly, both dimensions are governed by primacy and

interval (both temporal and pitch), suggesting that the fit judgments are driven by the match of early tones and early events, and larger intervals and longer durations. In addition, the fit judgments are not influenced by the surface-level matches defined by the specific coding of the exact match, the total number of match, or the number of match before the first mismatch in the surface structure. These specific types of correspondences do not predict the fit judgements, instead, the match of larger intervals and longer durations does.

DISCUSSION FOR EXPERIMENTS 4, 5, AND 6

Through a set of three experiments, we found the perception of both the bell and pentatonic patterns is dominated by surface-level features of primacy, and tone duration and interval size, respectively. The C minor tonal hierarchy and consonance measures also affected the perception of the pentatonic modes. In addition, the perceptual distances between the five rhythms and between the five modes do not show agreement with swap distance. Finally, the fit judgments revealed that the pitch-time parallel extends across dimensions, where they depended on the match of salience, not the surface-level match.

Two out of three other theoretical properties of the standard and diatonic patterns are shared by the bell and pentatonic patterns. The first is *maximally even set* (Clough & Douthett, 1991; Toussaint, 2013), where the 5 dark circles in Figures 5.1 and 5.2 cannot be more evenly distributed around the 12 clock positions or tone identities. The second characteristic is that the bell and pentatonic patterns are both *asymmetrical*, a characteristic that facilitates position-finding. The third characteristic of *deep rhythm* and *deep scale* does not hold for the bell and pentatonic patterns. As will be detailed in Chapter 8, the duration and interval size of “1” and “6” both occurred zero times, so zero is not unique among the distances between tone onsets or between pitches.

CHAPTER 8

GENERAL DISCUSSION

Chapter 8 considers the two sets of experiments, Experiments 1, 2, and 3 as compared to Experiments 4, 5, and 6. This chapter discusses shared cognitive principles and possible reasons behind the results that differed across the two sets of experiments. In addition, this chapter tests the subset relation between the bell and standard rhythms, and between the pentatonic and diatonic modes. The chapter ends with an overview of the complementary relation between the bell and standard rhythms, and between the pentatonic and diatonic modes.

Comparison between the Standard and Diatonic Patterns and the Bell and Pentatonic Patterns

The two sets of probe experiments (Experiments 1 and 2 versus Experiments 4 and 5) yielded a number of different results. For the rhythms, syncopation-shifted 3/2 metrical hierarchy and subgroupings governed the standard rhythmic pattern, whereas primacy and tone duration dominated the bell rhythmic pattern. For the modes, tone identity governed the diatonic scale pattern but primacy and interval size dominated the pentatonic scale pattern. Although tonal features underlay both mode patterns, tonality governed the diatonic scale pattern but the consonance measures were more prominent for the pentatonic scale pattern. In addition, swap distance predicted the perceptual arrangement of the seven standard rhythms and the seven diatonic modes, but not the five bell rhythms or the five pentatonic modes.

These differences revealed that primarily hierarchical-level principles governed the standard and diatonic patterns, but surface-level principles were more prominent for the bell and pentatonic patterns. A number of reasons might explain these differences. Listeners were less

familiar with the bell and pentatonic patterns than the standard and diatonic patterns¹. The lower familiarity suggests that listeners might not have enough exposure to the bell and pentatonic patterns to form hierarchical preferences. This explains the relative absence of top-down grouping for the bell rhythms, the weak effect of tonal hierarchies for the pentatonic modes, and the lack of organization for distances between rhythms and between modes.

There are fewer events and tones in total for the bell and pentatonic patterns (5) compared to the standard and diatonic patterns (7). Because of this, listeners received less information overall before rating the probe accents in Experiments 4 and 5. This might have increased the difficulty of the probe accent task.

Another reason is of a more theoretical nature, regarding *deep rhythm* (Toussaint, 2013) and *deep scale* (Gamer, 1967; Johnson, 2003; Toussaint, 2005). As mentioned at the end of Chapter 4, a pattern is *deep* when the distance between every pair of onsets or pitches occurs a unique number of times. Both the standard and diatonic patterns are deep, where each tone duration or interval size occurs a unique number of time. In contrast, the bell and pentatonic patterns do not have “1” as tone duration or as interval size in terms of semitone. This results in the distance of both “1” and “6” between any pair of onsets or pitches to occur zero times. Because the occurrence of zero is not unique to a specific distance, the bell and pentatonic patterns cannot be considered deep rhythms and deep modes, respectively. As a result, each event or tone is less unique in the bell and pentatonic patterns than in the standard and diatonic patterns. This lack of uniqueness might have caused confusion for listeners while they were attempting to organize individual events or tones into patterns.

¹ Experiment 1_Standard rhythm average = 32.81, Experiment 2_Diatonic mode average = 24.09, Experiment 3_Standard rhythm average = 33.92, Experiment 3_Diatonic mode average = 45.82; Experiment 4_Bell rhythm average = 30.11, Experiment 5_Pentatonic mode average = 23.68, Experiment 6_Bell rhythm average = 20.31, Experiment 6_Pentatonic mode average = 39.58.

Finally, Experiments 3 and Experiments 6 yielded similar results. The measure calculated by correlating the probe accent and the probe tone ratings predicted listeners' preference for how pitch and time combine perceptually for both sets of patterns. For the standard and diatonic patterns, findings revealed the preference for tonally stable tones to occur at rhythmically stable temporal positions. For the bell and pentatonic patterns, results showed the preference for early tones, tones that form a large interval with the following tone, and tones that are consonant with C to occur at early temporal positions and positions with long durations. Thus, the fit judgments uncovered a general preference for salient pitch-time co-occurrence. That is, the salient tones in a mode are preferred to occur at the salient events in a rhythm. This perceptual preference holds regardless of the specific salient features in each dimension that might vary across patterns. In addition, the three specific coding of surface-level match did not predict how pitch and time combine perceptually for either set of patterns. That is, combinations of rhythms and modes with matched durations and interval sizes were not preferred to other combinations.

The only difference in the interpretation of fit judgments between the two sets of patterns is that the standard and diatonic patterns perceptually interact, whereas the bell and pentatonic patterns do not. This is because both the standard and diatonic patterns are perceived hierarchically (rhythmic and tonal hierarchies), making their perceptual match an interaction between hierarchical levels. However, the bell and pentatonic patterns are driven by categorical features, such as primacy and interval. Thus, the match of salience based on these features simply shows preference for co-occurrence.

Subset Relation

The two sets of patterns, the standard and diatonic patterns as opposed to the bell and pentatonic patterns, are structurally related to each other in that one is the subset of the other.

Because the standard pattern was derived from the bell pattern (Pressing, 1983), the bell rhythms can be considered to be a subset of the standard rhythms. These two rhythms are also conceptualized as being the same (Jones, 1959, pp.212), and Yoruba players view the standard pattern as a form of the bell pattern, where splitting 3 into 2+1 seems to be a variation of the original 3. As Table 8.1 shows, by splitting 3 into 2+1, R1 of the bell pattern (2 2 3 2 3) becomes R4 of the standard pattern (2 2 2 1 2 2 1); by splitting 3 into 1+2, R1 of the bell pattern (2 2 3 2 3) becomes R5 of the standard pattern (2 2 1 2 2 1 2). Table 8.1 lists both types of split for each of the five bell rhythms. In addition, Pressing (1983) suggested a correspondence between R1 of the bell pattern (2 2 3 2 3) and R1 of the standard pattern (2 2 1 2 2 2 1), where the first 3 was split into 1+2 whereas the second 3 was split into 2+1. Because allowing both kinds of splits does not give unique mappings between the bell and standard rhythms, or between the pentatonic and diatonic modes, we will not consider them further. Regardless of the exact split, the two rhythmic patterns have a derivable relation to one another, such that the bell pattern is a subset of the standard pattern (Rahn, 1996). For example, all audible positions in Figure 5.1 fall on the audible positions in Figure 2.1.

To find out how these two rhythmic patterns are related to each other perceptually, we tested each of the two types of split, $3 = 2+1$ and $3 = 1 + 2$ (see Table 8.1). The 25 probe accent ratings of the bell pattern in Experiment 4 (see Chapter 5) were correlated with the 25 structurally corresponding probe accent ratings of the standard pattern in Experiment 1 (see Chapter 2). The two averaged probe accent ratings significantly correlated with each other for both the split of $3 = 2+1$ ($r(23) = .54, p = .0051$) and the split of $3 = 1+2$ ($r(23) = .65, p = .0005$). These correlations suggest the longer tone in the bell pattern (“3”) can be perceived as either 2+1 or 1+2, with the latter more prominent. However, because of the strong preference for accents on

the first tone, we eliminated it and then conducted the same analysis. The results were no longer significant. This suggests that the correlations were driven by this preference and not the perception of the bell pattern as a subset of the standard pattern. Thus, the theoretical and conceptual claims that the standard pattern was derived from and was considered as a form of the bell pattern was not supported by the perceptual data.

Similarly, for the modes, the pentatonic pattern is the diatonic pattern without the interval size of 1 semitone. The large interval size (3 semitones) can be broken into a semitone + a whole tone, or a whole tone + a semitone. Table 8.1 lists both types of split for each of the five

Table 8.1. *Perception of the bell rhythms as subsets of the standard rhythms, and the pentatonic scales as subsets of the diatonic scales.*

3 = 2+1		3 = 1+2	
Rhythm/Mode	Structure	Rhythm/Mode	Structure
Bell R1 / Pentatonic M1	2 2 3 2 3	Bell R1 / Pentatonic M1	2 2 3 2 3
Standard R4 / Diatonic M4	2 2 2 1 2 2 1	Standard R5 / Diatonic M5	2 2 1 2 2 1 2
Bell R2 / Pentatonic M2	2 3 2 3 2	Bell R2 / Pentatonic M2	2 3 2 3 2
Standard R5 / Diatonic M5	2 2 1 2 2 1 2	Standard R6 / Diatonic M6	2 1 2 2 1 2 2
Bell R3 / Pentatonic M3	3 2 3 2 2	Bell R3 / Pentatonic M3	3 2 3 2 2
Standard R6 / Diatonic M6	2 1 2 2 1 2 2	Standard R7 / Diatonic M7	1 2 2 1 2 2 2
Bell R4 / Pentatonic M4	2 3 2 2 3	Bell R4 / Pentatonic M4	2 3 2 2 3
Standard R1 / Diatonic M1	2 2 1 2 2 2 1	Standard R2 / Diatonic M2	2 1 2 2 2 1 2
Bell R5 / Pentatonic M5	3 2 2 3 2	Bell R5 / Pentatonic M5	3 2 2 3 2
Standard R2 / Diatonic M2	2 1 2 2 2 1 2	Standard R3 / Diatonic M3	1 2 2 2 1 2 2

pentatonic modes. Regardless of the exact split, the two mode patterns have a derivable relation to one another, such that the pentatonic pattern is a subset of the diatonic pattern. For example, all audible tones in Figure 5.2 fall on the audible positions in Figure 2.2.

To find out whether or not listeners perceived pitches in the pentatonic pattern similarly to the same pitches in the diatonic pattern, we tested each type of split. The 25 probe tone ratings of the pentatonic pattern in Experiment 5 (see Chapter 6) were correlated with the 25 structurally corresponding probe tone ratings of the diatonic pattern in Experiment 2 (see Chapter 3). The two probe tone ratings significantly correlated with each other for both the split of $3 = 2+1$ ($r(23) = .44, p = .0286$) and the split of $3 = 1+2$ ($r(23) = .69, p = .0002$). However, the correlations were no longer significant after eliminating the strong effect of the tone C. This suggests that the correlations were driven by C, not the perception of the pentatonic pattern as a subset of the diatonic pattern. Thus, the perceptual results do not reflect the subset relation for either the rhythms or the modes.

Complementary Relation

The bell and standard patterns are complementary rhythms, which was suggested to be a more fundamental relation compared to the subset relation (Pressing, 1983; Rahn, 1996; Royer & Garner, 1966; Toussaint, 2013). When two rhythms are complementary, the audible positions in one rhythm consists of the silent positions in the other rhythm, and vice versa. For example, when comparing Figure 2.1 and Figure 5.1, where the black circles mark audible positions and the white circles mark silent positions, the silent positions in Figure 2.1 form a pattern that is a rotated version of the audible positions in Figure 5.1. Together, the audible positions of the bell pattern and the standard pattern fill all 12 positions in the circle. Similarly, the pentatonic and diatonic patterns are complementary modes. The tones played in the pentatonic pattern (Figure 5.2) form a rotated version of silenced tones in the diatonic pattern (Figure 2.2). Together, the tones played in both patterns fill all 12 pitches in the circle. These complementary modes can be

thought of in terms of the 5 black keys and the 7 white keys within the same octave on the piano keyboard.

This complementary phenomenon also occurs in the domain of vision, and its perception is often referred to as figure-ground relationship (Peterson, Harvey, & Weidenbacher, 1991; Wagemans, Elder, Kubovy, Palmer, Peterson, & Singh, 2012). Examples such as the duck-rabbit image and the vase-face image showed that we see one part of the image as the figure and the other part as the ground. In addition, the figure and the ground can spontaneously switch roles, depending on which part of the image the border is seen as belonging to. The part of the image containing the border is the figure and the rest of the image is the ground.

In auditory perception, two principal analogies to the figure-ground relationship have been discussed. The first and most common analogy is with polyrhythms, where more than one rhythm is played concurrently. When two rhythms are presented at the same time, one rhythm tends to be perceived as the figure and the other as the ground. A rhythm is likely heard as the ground when its tempo is less than 500-700 ms but more than 200 ms, functioning like a meter (Handel & Oshinsky, 1981). Some experiments have found evidence that the “ground” rhythm bounds and sets off a “figure” rhythm, helping the listeners to form a coherent interpretation of the polyrhythm (Oshinsky & Handel, 1978). However, other evidence showed that the organizing principles of the “figure” rhythm determines the perceived goodness and difficulty of the polyrhythm (Garner, 1974, pp.62-63; Preusser, Garner, & Gottwald, 1970). Polyrhythms with more than 2 rhythms have also been studied (Magill & Pressing, 1997; Pressing, Summers, & Magill, 1996).

The second auditory analogy to the visual figure-ground relationship is with a single rhythm and considers the relationship between the sounded events and the silences (Beeman, 2006). That

these might be perceived as complementarity comes from considering how they are produced. The silent tones correspond to tensed muscles and lifted hands whereas the audible tones correspond to relaxed muscles and dropped hands (Rahn, 1996; von Hornbostel, 1928). Because of movements in producing music, some have argued that listeners would relate the sounds and silences to the drummers' movements, and hence consider the silences as equal and complementary to the sounded events.

The figure-ground relation of sound and silence is reflected well in our stimuli of a single rhythm with five and seven events in Experiments 1 and 4, respectively. However, because the probe accent technique requires the dynamically accented tone to be a sounded event, we did not probe the silent event. However, silent events have been probed in Palmer and Krumhansl (1990). Participants listened to a series of low-pitched beats and rated of how well the high-pitched beats that occurred every so often fit into the context beats. The ratings suggest that listeners mentally constructed metrical hierarchies without hearing a surface rhythm. Future studies of the complementary relation in rhythms can use this method to probe silent events in one rhythm, and compare the ratings with the probe accent ratings of sounded events in the complementary rhythm. Similarly, the figure-ground relation is also reflected in the five and seven tones in Experiments 2 and 5. However, the experiments focused on tones in the scales and did not probe tones that do not belong to the scale. Future studies can probe these "black keys" in a "white key" context.

Overall, this thesis addressed two fundamental questions in the perception and cognition of music. One question is how individual temporal events and pitches are perceived in the rhythmic and mode contexts, respectively. The other question is how pitch and time combine perceptually.

I investigated these questions by devising the novel probe accent technique, using the classic probe tone technique, and combining scale and rhythmic patterns.

Two sets of patterns were used, the standard and diatonic patterns and the bell and pentatonic patterns. They are ideal stimuli for empirically comparing the perception of pitch and time because these rhythms and modes are structurally isomorphic in tone duration and interval size. In addition, a number of music theories apply to both sets of patterns. Findings provided perceptual evidence for the theoretical proposals using the standard and diatonic patterns. Pitch-time interaction was found using stimuli combining the two patterns. However, most of these findings did not hold for the bell and pentatonic patterns despite the theoretical interest. Finally, I addressed alternative relations between the two sets of patterns, such as testing the subset relation. Further investigation of these patterns might shed light on the figure-ground relation in auditory perception.

References

- Agawu, K. (2006). Structural analysis of cultural analysis? Competing perspectives on the “Standard Pattern” of West African Rhythm. *Journal of the American Musicological Society*, 59, 1-46.
- Agmon, E. (1989). A mathematical model of the diatonic system. *Journal of Music Theory*, 33, 1-25.
- Agmon, E. (1997). Musical durations as mathematical intervals: Some implications for the theory and analysis of rhythm. *Musical Analysis*, 16, 45-75.
- Anku, W. (2000). Circles and time: A theory of structural organization of rhythm in African music. *Music Theory Online*, 6.
- Beeman, W. O. (2006). Silence in Music. M.-L. Achino-Loeb (Eds.), *Silence: The Currency of Power* (pp. 23-34). New York, NY: Berghahn Books.
- Boltz, M. G. (1998). Tempo discrimination of musical patterns: Effects due to pitch and rhythmic structure. *Perception & Psychophysics*, 60, 1357-1373.
- Boltz, M. G. (1995). Effects of event structure on retrospective duration judgments. *Perception & Psychophysics*, 57, 1080-1096.
- Boltz, M. G. (1998). Tempo discrimination of musical patterns: Effects due to pitch and rhythmic structure. *Perception & Psychophysics*, 60, 1357-1373.
- Boltz, M. G. (1999). The processing of melodic and temporal information: Independent or unified dimensions? *Journal of New Music Research*, 28, 67-79.
- Bowling, D. L., Purves, D., & Gill, K. Z. (2018). Vocal similarity predicts the relative attraction of musical chords. *PNAS*, 115, 216-221.

- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- Brochard, R., Abecasis, D., Potter, D., Ragot, R., & Drake, C. (2003). The “ticktock” of our internal clock: Direct brain evidence of subjective accents in isochronous sequences. *Psychological Science*, 14, 362-366.
- Browne, R. (1981). Tonal implications of the diatonic set. *In Theory Only*, 5, 3-21.
- Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L. (1984). Tonal hierarchies in the music of North India. *Journal of Experimental Psychology: General*, 113, 394-412.
- Chernoff, J. M. (1979). *African rhythm and African sensibility: Aesthetics and social action in African musical idioms*. Chicago, IL: The University of Chicago Press.
- Clarke, E. F. (1987). Categorical rhythm perception: An ecological perspective. In A. Gabrielsson (Eds.), *Action and perception in rhythm and music* (pp.19-33). Stockholm: Royal Swedish Academy of Music.
- Clough, J., & Douthett, J. (1991). Maximally Even Sets. *Journal of Music Theory*, 35, 93-173.
- Cooper, G. W., & Meyer, L. B. (1960). *The rhythmic structure of music*. Chicago: The University of Chicago Press.
- Curtis, L. (1992). Mode. T. Knighton & D. Fallows (Eds.), *Companion to Medieval and Renaissance Music* (pp. 255-264). New York, NY: Schirmer Books.
- Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. *Perception & Psychophysics*, 28, 381-389.
- Deutsch, D. (1995). *Musical illusions and paradoxes* [CD]. La Jolla, CA: Philomel Records.
- Gamer, C. (1967). Deep scales and difference sets in equal-tempered systems. *In Proceedings of the Second Annual Conference of the American Society of University Composers*, 113-122.

- Garner, W. (1974). *The processing of information and structure*. New York, NY: John Wiley and Sons, 1974.
- Garner, W. R. (1976). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology*, 8, 98-123.
- Gill, K. Z., & Purves, D. (2009). A biological rationale for musical scales. *PLoS ONE*, 4, e8144. 10.1371/journal.pone.0008144
- Handel, S. (1974). Perceiving melodic and rhythmic auditory patterns. *Journal of Experimental Psychology*, 103, 922-933.
- Handel, S., & Oshinsky, J. S. (1981). The meter of syncopated auditory polyrhythms. *Perception & Psychophysics*, 30, 1-9.
- Handel, S., & Todd, P. (1981). Segmentation of sequential patterns. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 41-55.
- Hannon, E. E., Snyder, J. S., Eerola, T., & Krumhansl, C. L. (2004). The role of melodic and temporal cues in perceiving musical meter. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 956-974.
- Helmholtz, H. L. F. (1954). *On the sensations of tone as a physiological basis for the theory of music*. A. J. Ellis (Eds. & trans.). New York: Dover. (Revised edition originally published, 1885.)
- Hébert, S., & Peretz, I. (1997). Recognition of music in long-term memory: Are melodic and temporal patterns equal partners? *Memory & Cognition*, 25, 518-533.
- Huron, D., & Veltman, J. (2006). A cognitive approach to Medieval mode: Evidence for an historical antecedent to the major/minor system. *Empirical Musicology Review*, 1, 33-55.
- Hutchinson, W., & Knopoff, L. (1978). The acoustic component of Western consonance.

Interface, 7, 1-29.

Järvinen, T. (1995). Tonal hierarchies in jazz improvisation. *Music Perception*, 12, 415-437.

Järvinen, T., & Toiviainen, P. (2000). The effect of metre on the use of tones in jazz improvisation. *Musicae Scientiae*, 4, 55-74.

Johnson, T. A. (2003). Foundations of diatonic theory: A mathematically based approach to music fundamentals. Emeryville, CA: Key College.

Jones, A. M. (1958). On transcribing African music. *African Music*, 2, 11-14.

Jones, A. M. (1959). *Studies in African Music*. London: Oxford University Press.

Jones, M. R. (1993). Dynamics of musical patterns: How do melody and rhythm fit together? In T. J. Tighe & W. J. Dowling (Eds.), *Psychology and Music: The Understanding of Melody and Rhythm* (pp. 67-92). Hillsdale, NJ: Lawrence Erlbaum Associates.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, 96, 459-491.

Jones, M. R., Boltz, M., & Kidd, G. (1982). Controlled attending as a function of melodic and temporal context. *Perception & Psychophysics*, 32, 211-218.

Jones, M. R., Summerell, L., & Marshburn, E. (1987). Recognizing melodies: A dynamic interpretation. *The Quarterly Journal of Experimental Psychology*, 39, 89-121.

Kameoka, A., & Kuriyagawa, M. (1969). Consonance theory Part II: Consonance of complex tones and its calculation method. *Journal of the Acoustical Society of America*, 45, 1460-69.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.

- Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin*, 126, 159-179.
- Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, 89, 334-368.
- Krumhansl, C. L., Sandell, G. J., & Sergeant, D. C. (1987). The perception of tone hierarchies and mirror forms in twelve-tone serial music. *Music Perception*, 5, 31-78.
- Krumhansl, C. L., & Shepard, R. N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 579-594.
- Kubik, G. (1999). A strange absence. In *Africa and the Blues* (pp. 51-62). Jackson, MS: University Press of Mississippi.
- Lading, O., Honing, H., Haden, G., & Winkler, I. (2009). Probing attentive and preattentive emergent meter in adult listeners without extensive musical training. *Music Perception*, 26, 377-386.
- Large, E. W. (2010). A dynamical systems approach to musical tonality. In R. Huys & V. Jirsa (Eds.), *Nonlinear dynamics in human behavior* (pp. 193-211). New York: Springer.
- Large, E. W., Kim, J. C., Flaig, N. K., Bharucha, J. J., & Krumhansl, C. L. (2016). A neurodynamic account of musical tonality. *Music Perception*, 33, 319-331.
- Large, E. W., & Kolen, J. F. (1994). Resonance and the perception of musical meter. *Connection Science*, 6, 177-208.
- Lelo-de-Larrea-Mancera, E. S., Rodríguez-Agudelo, Y., & Solís-Vivanco, R. (2017). Musical rhythm and pitch: A differential effect on auditory dynamics as revealed by the N1/MMN/P3a complex. *Neuropsychologia*, 100, 44-50.

- Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: The MIT Press.
- Locke, D. (1982). Principles of offbeat timing and cross-rhythm in southern Ewe dance drumming. *Ethnomusicology*, 26, 217-246.
- London, J. (2002). Some non-isomorphisms between pitch and time. *Journal of Music Theory*, 46, 127-151.
- London, J. (2004). *Hearing in time: Psychological aspects of musical meter*. New York: Oxford University Press.
- Magill, J., & Pressing, J. L. (1997). Asymmetric cognitive clock structures in West African Rhythms. *Music Perception*, 15, 189-222.
- Malmberg, C. F. (1918). The perception of consonance and dissonance. *Psychological Monographs*, 25, 93-133.
- Martens, P., & Benadon, F. (2017). Musical structure: Time and rhythm. In R. Ashley & R. Timmers (Eds.), *The Routledge Companion to Music Cognition* (pp. 115-127). New York, NY: Routledge.
- McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2010). Individual differences reveal the basis of consonance. *Current Biology*, 20, 1035-1041.
- Monahan, C. B., & Carterette, E. C. (1985). Pitch and duration as determinants of musical space. *Music Perception*, 3, 1-32.
- Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model*. Chicago, IL: University of Chicago Press.
- Nketia, J. H. K. (1974). *The music of Africa*. New York, NY: W. W. Norton & Company Inc.

- Oram, N., & Cuddy, L. L. (1995). Responsiveness of Western adults to pitch-distributional information in melodic sequences. *Psychological Research*, 57, 103-118.
- Oshinsky, J. S., & Handel, S. (1978). Syncopated auditory polyrhythms: Discontinuous reversals in meter interpretation. *The Journal of the Acoustical Society of America*, 63, 936-939.
- Pressing, J., Summers, J., & Magill, J. (1996). Cognitive multiplicity in polyrhythmic pattern performance. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 1127-1148.
- Preusser, D., Garner, W. R., & Gottwald, R. L. (1970). Perceptual organization of two-element temporal patterns as a function of their component one-element patterns. *The American Journal of Psychology*, 83, 151-170.
- Palmer, C., & Krumhansl, C. L. (1987a). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 116-126.
- Palmer, C., & Krumhansl, C. L. (1987b). Pitch and temporal contributions to musical phrase perception: Effects of harmony, performance timing, and familiarity. *Perception & Psychophysics*, 41, 505-518.
- Palmer, C., & Krumhansl, C. L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 728-741.
- Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, 6, 688-691.
- Peretz, I. & Kolinsky, R. (1993). Boundaries of separability between melody and rhythm in music discrimination: A neuropsychological perspective. *The Quarterly Journal of Experimental Psychology*, 46A, 301-325.

- Peretz, I., & Morais, J. (1989). Music and modularity. *Contemporary Music Review*, 4, 279-293.
- Peterson, M. A., Harvey, E. M., & Weidenbacher, H. J. (1991). Shape recognition contributions to figure-ground reversal: Which route counts? *Journal of Experimental Psychology: Human Perception and Performance*, 17, 1075-1089.
- Powers, H. (2001). Mode. In Grove Music Online.
- Pressing, J. (1983). Cognitive isomorphisms between pitch and rhythm in world musics: West Africa, the Balkans and Western Tonality. *Studies in Music*, 17, 38-61.
- Pressing, J., Summers, J., & Magill, J. (1996). Cognitive multiplicity in polyrhythmic pattern performance. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 1127-1148.
- Preusser, D., Garner, W. R., & Gottwald, R. L. (1970). Perceptual organization of two-element temporal patterns as a function of their component on-element patterns. *The American Journal of Psychology*, 83, 151-170.
- Prince, J. B. (2011). The integration of stimulus dimensions in the perception of music. *The Quarterly Journal of Experimental Psychology*, 64, 2125-2152.
- Prince, J. B. (2014). Contributions of pitch contour, tonality, rhythm, and meter to melodic similarity. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 2319-2337.
- Prince, J. B., & Loo, L-M. (2017). Surface and structural effects of pitch and time on global melodic expectancies. *Psychological Research*, 81, 255-270.
- Prince, J. B., & Pfordresher, P. Q. (2012). The role of pitch and temporal diversity in the perception and production of musical sequences. *Acta Psychologica*, 141, 184-198.

- Prince, J. B., & Schmuckler, M. A. (2014). The tonal-metric hierarchy: A corpus analysis. *Music Perception, 31*, 254-270.
- Prince, J. B., Schmuckler, M. A., & Thompson, W. F. (2009). The effect of task and pitch structure on pitch-time interactions in music. *Memory & Cognition, 37*, 368-381.
- Rahn, J. (1987). Asymmetrical ostinatos in sub-Saharan music: Time, pitch, and cycles reconsidered. *In Theory Only, 9*, no.7: 23-37.
- Rahn, J. (1996). Turning the analysis around: Africa-derived rhythms and Europe-derived music theory. *Black Music Research Journal, 16*, 71-89.
- Razdan, A. S., & Patel, A. D. (2016). Rhythmic consonance and dissonance: Perceptual ratings of rhythmic analogs of musical pitch intervals and chords. *In Proceedings of the International Conference on Music Perception and Cognition*, 807-812.
- Rechberger, H. (2008). *Scales and modes around the world*. Helsinki, Finland: Fennica-Gehrman.
- Peterson, M. A., Harvey, E. M., & Weidenbacher, H. J. (1991). Shape recognition contributions to figure-ground reversal: Which route counts? *Journal of Experimental Psychology: Human Perception and Performance, 17*, 1075-1089.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*, 192-233.
- Rosch, E., & Mervis, C. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573-605.
- Rosenthal, M. A., & Hannon, E. E. (2016). Cues to perceiving tonal stability in music: The role of temporal structure. *Music Perception, 33*, 606-612.

- Royer, F. L., & Garner, W. R. (1966). Response uncertainty and perceptual difficulty of auditory temporal patterns. *Perception & Psychophysics*, *1*, 41-47.
- Stolzenburg, F. (2015). Harmony perception by periodicity detection. *Journal of Mathematics and Music*, *9*, 215-238.
- Temperley, D. (1999). Syncopation in Rock: A perceptual perspective. *Popular Music*, *18*, 19-40.
- Temperley, D. (2000). Meter and grouping in African music: A review from music theory. *Ethnomusicology*, *44*, 65-96.
- Temperley, D. (2018). *The musical language of rock*. New York, NY: Oxford University Press.
- Temperley, D., & Tan, D. (2013). Emotional connotations of diatonic modes. *Music Perception*, *30*, 237-257.
- Tillmann, B., & Lebrun-Guillaud, G. (2006). Influence of tonal and temporal expectations on chord processing and on completion judgments of chord sequences. *Psychological Research*, *70*, 345-358.
- Toussaint, G. T. (2005). The geometry of musical rhythm. In *Proceedings of the Japanese Conference on Discrete and Computational Geometry*, 198-212. Berlin & Heidelberg, Germany: Springer.
- Toussaint, G. T. (2013). *The geometry of musical rhythm: What makes a “good” rhythm good?* Boca Raton, FL: CRC Press.
- von Hornbostel, E. M. (1928). African negro music. *Africa: Journal of the International African Institute*, *1*, 30-62.
- Vuust, P., & Witek, M. A. G. (2014). Rhythmic complexity and predictive coding: A novel approach to modeling rhythm and meter perception in music. *Frontiers in Psychology*, *5*, 1-14.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., & Singh, M. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, 138, 1172-1217.